Évaluer les apprentissages

Les apprentissages de la psychologie cognitive

Jacques Grégoire (Éd.)







La place de plus en plus grande accordée aujourd'hui à l'**évaluation formative** et à l'**évaluation diagnostique** modifie en profondeur les besoins des praticiens. Dans ce cadre, l'important est moins d'enregistrer la réussite ou l'échec que de comprendre ce qui est sous-jacent aux performances observées. Quelles **démarches** l'élève a-t-il suivies pour arriver à la réponse qu'il nous propose? A-t-il compris le **sens de la procédure** utilisée? Où a-t-il rencontré des **difficultés?** Comment a-t-il fait face à ces difficultés? Répondre à ces questions est essentiel pour pouvoir intervenir de manière efficiente.

Des réponses appropriées se trouvent dans cet ouvrage, fruit d'une collaboration de spécialistes de la psychologie cognitive et de l'évaluation.
L'ensemble est organisé en quatre grands thèmes:
1. l'évaluation des apprentissages mathématiques;
2. l'évaluation de la lecture;
3. la place de la métacognition dans l'évaluation;
4. les implications psychométriques des nouveaux modèles d'évaluation.

Avec les contributions de:
J. Bideaud, *Université Charles de Gaulle-Lille 3.*
J.-P. Fischer, *IUFM de Lorraine.*
J. Grégoire, *Université Catholique de Louvain, Louvain-la Neuve.*
C. Jourdain, *Université de Bourgogne, Dijon.*
D. Laveault, *Université d'Ottawa.*
B. Lété, *INRP, Paris.*
J. Retschitzki, *Université de Fribourg.*
G. Scallon, *Université Laval, Québec.*
J.-L. Wolfs, *Université Libre de Bruxelles.*
D. Zagar, *Université de Bourgogne, Dijon.*

ISBN : 2-8041-2237-9





Collection dirigée par
Philippe JONNAERT, Sherbrooke

Comité scientifique international
Dan Baba TAHIROU, *Niamey*; Jean-Marie DE KETELE, *Louvain-la-Neuve;* Maurice SACHOT, *Strasbourg;* Jacques TARDIF, *Sherbrooke.*

Animée par Philippe Jonnaert (Université de Sherbrooke), voici une collection en sciences de l'éducation créée pour soumettre à la critique des praticiens les réflexions théoriques et les résultats de recherches et de travaux actuels et pour offrir aux enseignants et aux professionnels de l'éducation des outils pour leur pratique quotidienne et une réflexion sur ces derniers.

BONAMI M. et GARANT M. (Éds), *Systèmes scolaires et pilotage de l'innovation. Émergences et implantation du changement.*
BOURGEOIS É. (Éd.), *L'adulte en formation : regards pluriels.*
CAILLOT M. et RAISKY C. (Éds), *Le didactique au-delà des didactiques. Regards croisés sur des concepts fédérateurs.*
DONNAY J. et ROMAINVILLE M. (Éds), *Enseigner à l'Université. Un métier qui s'apprend?*
FRANCQ B. et MAROY C. (Éds), *Formation et socialisation au travail.*
GRÉGOIRE J. (Éd.), *Évaluer les apprentissages. Les apports de la psychologie cognitive.*
LEPOT-FROMENT C. (Éd.), *Éducation spécialisée. Recherches et pistes d'action.*
MATHY C. (Éd.), *Enseignement des sciences et valeurs.*
PAQUAY L., ALTET M., CHARLIER E. et PERRENOUD P. (Éds), *Former des enseignants professionnels. Quelles stratégies? Quelles compétences?*

Évaluer les apprentissages

Les apports de la psychologie cognitive

Jacques Grégoire (Éd.)





Département De Boeck Université
Paris, Bruxelles

Printed in Belgium

D 1996/0074/93 ISBN 2-8041-2237-9

Sommaire

Table des matières

CHAPITRE 4

La mesure des Temps de Réponses aux calculs numériques élémentaires en situation scolaire 81

DEUXIÈME PARTIE

LA LECTURE

CHAPITRE 5

Introduction

Jacques Grégoire

L'évaluation des apprentissages scolaires reste fortement influencée par un point de vue behavioriste. En effet, l'attention des enseignants et des psychologues scolaires se concentre souvent sur une réalité observable de manière directe et objective : la performance de l'élève. Ce point de vue peut être justifié lorsque le but de l'évaluation est de sélectionner, de classer ou de certifier. Par exemple, lors d'un examen d'admission, la quantification des performances et leur comparaison à un seuil minimum procure généralement des informations suffisantes pour pouvoir prendre des décisions adéquates.

La place de plus en plus grande accordée aujourd'hui à l'évaluation formative et à l'évaluation diagnostique modifie en profondeur les besoins des praticiens. Dans ce cadre, l'important est moins d'enregistrer la réussite ou l'échec que de comprendre ce qui est sous-jacent aux performances observées. Quelle démarche l'élève a-t-il suivi pour arriver à la réponse qu'il nous propose ? A-t-il compris le sens de la procédure utilisée ? Où a-t-il rencontré des difficultés ? Comment a-t-il fait face à ces difficultés ? Répondre à ces questions est essentiel pour pouvoir intervenir de manière efficiente. En fait, dans une perspective formative ou remédiative, la performance observée doit être prise comme un symptôme de l'apprentissage ou de la difficulté d'apprentissage. Le travail de l'évaluateur est de décoder ce symptôme, de découvrir sa véritable signification. Ici, l'erreur n'est plus vue comme une faute mais comme un indicateur du dysfonctionnement d'un apprentissage. Son analyse est le point de départ obligé de toute démarche d'aide.

Comprendre ce qui se passe dans la tête de l'élève implique que le praticien réalise, sur base de ses observations, des inférences à propos d'une réalité qu'il ne peut directement observer. Pour pouvoir réaliser ces inférences, il a besoin d'une théorie du fonctionnement mental et de l'apprentissage. De ce point de vue, la théorie piagétienne a constitué, et constitue encore, une référence importante pour de nombreux enseignants et psychologues scolaires. Elle offre en effet un cadre conceptuel pour comprendre la construction progressive du raisonnement de l'enfant. Elle permet en particulier d'expliquer certaines difficultés d'apprentissage par la non acquisition d'opérations mentales relativement générales. Malheureusement, la théorie piagétienne reste souvent insuffisante pour comprendre de manière plus précise les apprentissages en situation scolaire. En effet, elle ne prend guère en compte les facteurs contextuels et proprement psychologiques qui, inévitablement, interviennent dans tout apprentissage.

La psychologie cognitive apparaît aujourd'hui comme une référence théorique importante pour le praticien soucieux de comprendre le sujet apprenant. Elle lui procure de nouveaux concepts qui viennent combler les lacunes de la théorie piagétienne. Toutefois, à l'inverse de la théorie piagétienne, la psychologie cognitive ne constitue pas une théorie unifiée. Il serait donc plus correct de parler des théories cognitives, au pluriel.

Bien que la transition entre le behaviorisme et le cognitivisme se soit faite progressivement, on peut situer le démarrage du courant cognitiviste autour de 1960. Avec le recul, la conférence donnée par Allen Newell et Her-

bert Simon en 1958 (Lachman et al., 1979) semble avoir joué le rôle d'un déclencheur. Newell et Simon proposèrent alors une analogie entre le traitement de l'information par l'ordinateur et par l'être humain. La nouveauté ne provenait pas seulement du rapprochement entre la pensée humaine et le fonctionnement de l'ordinateur. Elle venait surtout de l'affirmation d'un point de vue radicalement différent de celui du behaviorisme. En effet, les chercheurs portaient leur attention sur ce qui se passait entre le stimulus et la réponse et prétendaient étudier les processus internes de manière scientifique.

L'affirmation que les processus mentaux existent et peuvent être étudiés rigoureusement constitue le premier postulat sur lequel s'accordent les chercheurs qui s'inscrivent dans le courant cognitiviste. Un second postulat est que l'être humain traite de manière active l'information reçue : il sélectionne, met en relation, calcule... Enfin un troisième postulat sur lequel se rejoignent les cognitivistes est qu'il est possible d'évaluer ce traitement sur la base du temps de réponse et de la précision de la réponse. En effet, les traitements mentaux prennent du temps. Par conséquent, il est possible d'estimer quels processus mentaux sont mis en oeuvre, en comparant le temps mis pour répondre à différents stimuli. Dans le présent ouvrage, le chapitre de Fischer propose une intéressante illustration de l'usage du temps de réponse pour l'étude des processus mentaux. La précision de la réponse est également un bon indicateur des traitements mentaux réalisés par un sujet. Pour pouvoir répondre correctement à certains stimuli, le sujet peut en effet être contraint d'utiliser une procédure déterminée. Les erreurs qu'il va éventuellement commettre seront alors la manifestation de l'usage de procédures alternatives, moins adéquates.

Il est évident que les théories cognitives offrent un cadre conceptuel nous permettant de mieux comprendre les apprentissages des élèves. Mais le passage de la théorie à la pratique ne va malheureusement pas de soi. La majorité des chercheurs en psychologie cognitive ne se sont pas intéressés aux applications de leurs découvertes. Il existe donc un important décalage temporel entre le développement des modèles théoriques et leur traduction dans les pratiques d'évaluation. Les colloques et les publications sur cette question commencent seulement à apparaître. Et de nombreuses questions théoriques et méthodologiques restent posées.

C'est dans ce contexte que, dans le cadre des rencontres du Réseau Education et Formation (REF) de Louvain-la-Neuve (septembre 1994), nous avons choisi de consacrer un symposium à la question des apports de la psychologie cognitive à l'évaluation des apprentissages scolaires. Durant deux jours, des spécialistes de la psychologie cognitive et de l'évaluation des apprentissages se sont rencontrés pour échanger sur cette question. Le présent ouvrage est le fruit de ces rencontres particulièrement stimulantes pour la réflexion. Nous espérons que le lecteur y trouvera une source d'inspiration et un ferment pour le développement de ses propres pratiques.

Pour faciliter la consultation de cet ouvrage, nous avons regroupé les communications selon quatre grands thèmes : (1) l'évaluation des apprentissages mathématiques; (2) l'évaluation de la lecture; (3) la place de la métaco-

gnition dans l'évaluation et (4) les implications psychométriques des nouveaux modèles d'évaluation.

La première partie, consacrée à la mathématique, comprend quatre articles. Dans le premier de ceux-ci, *Jacques Grégoire* réfléchit sur les modèles utilisés pour l'évaluation diagnostique des troubles des apprentissages mathématiques. Il confronte en particulier le modèle piagétien et différents modèles cognitivistes. *Jacqueline Bideaud* aborde ensuite la question de l'évaluation des capacités "mathématiques" que possèdent déjà les jeunes enfants avant d'aborder l'apprentissage de la mathématique élémentaire. Elle brosse par la même occasion un tableau des théories piagétiennes et néopiagétiennes de l'acquisition de la notion de nombre. *Jean Retschitzki* présente, dans une perpective historique, les apports de la psychologie cognitive pour la compréhension de la résolution de problèmes. Il en souligne certaines limites et note l'émergence de modèles explicatifs concurrents. Enfin, *Jean-Paul Fischer* décrit la mise en place en situation scolaire d'une évaluation du temps de réponse à des calculs élémentaires. Les possibilités diagnostiques de cette méthode d'évaluation sont illustrées et discutées.

La seconde partie de l'ouvrage est centrée sur le thème de la lecture. Le texte de *Christine Jourdain, Daniel Zagar et Bernard Lété* est consacré à l'évaluation diagnostique des difficultés de lecture chez les adultes. Un cadre théorique est proposé et un ensemble d'épreuves est présenté, permettant une évaluation pas à pas des processus en jeu dans l'acte de lecture. *Bernard Lété* présente l'usage que l'on peut faire de l'ordinateur pour améliorer les capacités en lecture. Au travers d'une analyse minutieuse de la littérature sur cette question, il nous montre comment, en s'appuyant sur un modèle de la lecture, l'ordinateur peut devenir un outil d'évaluation formative particulièrement efficace.

La troisième partie de l'ouvrage concerne l'utilisation du concept de métacognition dans le cadre de l'évaluation formative. *Gérard Scallon* nous retrace l'évolution des pratiques d'évaluation formative et analyse les difficultés rencontrées. Il nous montre comment la psychologie cognitive, particulièrement grâce au concept de métacognition, a permis à l'évaluation formative de sortir des ornières dans lesquelles elle risquait de s'enliser. *José-Luis Wolfs* développe plus spécifiquement l'usage de la métacognition comme moyen de faire participer l'apprenant à l'évaluation diagnostique de ses apprentissages. Il présente une grille permettant d'analyser les pratiques pédagogiques qui vont dans ce sens.

Enfin, l'ouvrage se clôture par le texte de *Dany Laveault* qui est consacré à la question des relations entre les modèles cognitifs et les méthodes psychométriques. Il nous montre que la théorie de la généralisabilité peut être un outil puissant pour nous aider à mesurer de manière valide et fiable les différentes facettes du traitement d'un problème par un sujet.

1e PARTIE

LA MATHÉMATIQUE



Quelle démarche d'évaluation diagnostique des troubles d'apprentissage en mathématique ?

Jacques Grégoire

Introduction

Lorsque l'on compare l'évaluation de la lecture et celle des apprentissages mathématiques, on constate un écart considérable quant au nombre de travaux et de recherches qui y sont consacrés. L'impact des travaux piagétiens sur le nombre ont quelque peu masqué le manque d'investissement du champ des mathématiques par les spécialistes de l'évaluation. Il a fallu attendre le début des années 80 et les recherches cognitivistes sur les mathématiques pour voir cet investissement quelque peu grandir. Mais, comme nous le verrons, les carences restent énormes et les praticiens désireux de réaliser une évaluation diagnostique des apprentissages mathématiques souffrent du manque d'outils et de référentiels théoriques suffisamment larges et solides.

Le présent texte est consacré à une analyse critique des modèles théoriques qui peuvent guider l'évaluation diagnostique des apprentissages mathématiques. Nous discutons de leurs postulats de base ainsi que de leurs avantages et limites. Nous terminons en proposant une intégration des différents points de vue au sein de la démarche diagnostique.

1. QU'EST-CE QUE L'ÉVALUATION DIAGNOSTIQUE DES APPRENTISSAGES ?

Le concept d'évaluation diagnostique n'est pas défini de manière uniforme par tous les spécialistes. Il soulève par ailleurs des questions de fond qui, étonnamment, sont rarement abordées et discutées. Dans cette section, nous commencerons par préciser ce que nous entendons par *évaluation diagnostique*. Nous expliciterons et discuterons ensuite plusieurs postulats sur lesquels s'appuie la démarche d'évaluation diagnostique.

Se référant à la typologie des formes d'évaluation proposées par Bloom et ses collaborateurs, Scallon (1988a) souligne que, pour ces auteurs, l'évaluation diagnostique remplit deux fonctions. La première est de nature préventive et concerne l'intégration des élèves dans une nouvelle séquence d'apprentissage. Dans ce cas, l'évaluation diagnostique vise à mettre en évidence les forces et les faiblesses de chaque élève afin de préciser le point d'entrée adéquat dans la séquence d'apprentissage et de déterminer le mode d'enseignement le plus adapté.

Une seconde fonction de l'évaluation diagnostique est de «*déterminer la cause de difficultés persistantes chez certains élèves*» (Scallon, 1988a, p.69). Sous certains aspects, cette fonction rejoint celle de l'évaluation formative qui a également pour but de déceler les difficultés pouvant se présenter en cours d'apprentissage afin d'y remédier rapidement. Par rapport à l'évaluation formative, la spécificité de l'évaluation diagnostique est d'être à la fois plus approfondie et plus globale (prise en compte des facteurs motivationnels, environnementaux...). Mais les limites restent floues. Scallon propose une distinction plus nette en se référant aux types de cause des dif-

ficultés d'apprentissage. L'évaluation diagnostique s'intéresserait aux causes exogènes à la situation d'apprentissage alors que l'évaluation formative prendrait uniquement en compte les causes endogènes à cette même situation. En ce sens, le caractère diagnostique de l'évaluation formative resterait spécifiquement pédagogique.

De son côté, Cardinet (1988) ne retient pas cette distinction et parle d'*évaluation formative diagnostique* pour désigner tout examen approfondi des difficultés des élèves quelles qu'en soient les causes. Par ailleurs, il évite d'utiliser le terme de diagnostique à propos de ce qu'il appelle l'*évaluation formative de départ*, qui désigne l'appréciation des pré-acquis des élèves dans le domaine à étudier. Pour notre part, nous adoptons cette distinction qui permet d'éviter les confusions et nous réservons le terme d'évaluation diagnostique pour désigner l'examen approfondi des troubles d'apprentissages en vue d'entreprendre une action de remédiation. Cette évaluation peut se faire au sein de la classe ou en dehors de celle-ci. Dans ce dernier cas, elle peut être menée par des psychologues scolaires, des logopèdes...

Mais qu'est-ce qu'une évaluation approfondie des troubles d'apprentissage ? Cette démarche ne va pas de soi et plusieurs de ses postulats, souvent implicites, demandent à être discutés. Le premier concerne le niveau de la réalité à évaluer. Depuis les travaux de Chomsky[1], la distinction entre compétence et performance a été largement adoptée. La performance fait référence aux comportements observables produits par le sujet. Quant à la compétence, elle est constituée d'un ensemble de capacités organisées qui sous-tendent les performances. Les compétences ne sont pas directement observables mais peuvent seulement être inférées à partir des performances.

Au niveau de la démarche diagnostique, nous retrouvons cette distinction. Certaines démarches sont en effet centrées sur la tâche et la performance; elles sont fondamentalement a-théoriques. D'autres, par contre, s'appuient sur un modèle théorique de la réalité sous-jacente aux performances. De ce choix initial découle un certain nombre de contraintes concernant les outils d'évaluation. Dans le cas d'une approche centrée sur les compétences, il est nécessaire de valider la liaison entre la compétence et la tâche prise comme son révélateur. En particulier, il faut éviter que les tâches sélectionnées ne mesurent des capacités étrangères aux compétences visées. Dans le cas d'une approche centrée sur les performances, ce dernier problème ne se pose guère. En effet, dans la mesure où l'on ne s'intéresse qu'aux comportements observables, la nature exacte des capacités sous-jacentes n'est pas prise en compte. Par contre, on est en droit de s'interroger sur la possibilité de

---

1 *«Nous faisons une distinction fondamentale entre la compétence (la connaissance qu'a le locuteur-auditeur de sa propre langue) et la performance (l'usage effectif du langage dans des situations concrètes). [...] Le problème pour le linguiste, comme pour l'enfant qui apprend un langage, est de déterminer à partir de données de performance le système de règles sous-jacent qui a été maîtrisé par le locuteur-auditeur et que celui-ci utilise dans sa performance effective»* (Chomsky, p.4, 1965).

généraliser le comportement observé. Ici, le résultat risque en effet d'être très dépendant de la tâche proposée.

Un second postulat qui sous-tend l'évaluation diagnostique des apprentissages concerne la place que l'on attribue au contexte. Certains auteurs (Resnick & Resnick, 1991) soulignent que les capacités des sujets prennent une forme différente selon les contextes où elles se révèlent. Pour ceux-ci, une évaluation valide ne peut se concevoir que contextualisée. Pour être correctement évalué, le sujet doit en effet être confronté à une tâche signifiante et motivante. Cet accent mis sur l'importance du contexte est certes pertinent mais il est aussi simplificateur. Un même contexte n'est en effet pas également stimulant et facilitant pour tous les sujets, ce qui pose des problèmes d'équité et de généralisation des résultats. Par ailleurs, il est incorrect d'affirmer qu'il y aurait des tests contextualisés et d'autres pas. Comme le fait remarquer Sternberg (1990, p.221), même dans les tests standardisés *«les items sont contextualisés en relation avec l'environnement dans lequel ils sont supposés mesurer les performances, c'est-à-dire l'environnement scolaire».* Par conséquent, l'opposition n'est pas entre les mesures contextualisées et les mesures décontextualisées mais entre les premières et les mesures *«transcontextualisées»* (Messick, 1994). Messick souligne en effet que, si l'on veut éviter d'être confronté à une prolifération de capacités qualitativement différentes selon le contexte d'usage, nous devons tirer des inférences en prenant en compte la consistance des performances au travers des contextes. Cette consistance est alors prise comme un indicateur fiable des compétences visées.

Enfin, un troisième postulat de l'évaluation diagnostique est lié à la légitimité que l'on reconnaît ou non à la décomposition des compétences complexes en aptitudes élémentaires. Resnick et Resnick (1990) plaident pour une évaluation qui respecte la complexité des capacités car une évaluation qui se centre sur les aptitudes élémentaires ne prend pas en compte les interactions entre celles-ci. Et il est vrai que les compétences ne sont pas égales à la somme des aptitudes qui les composent. Toutefois, on imagine mal comment un outil diagnostique pourrait faire l'économie d'une analyse des compétences complexes. De plus, on peut douter de la possibilité d'apprécier véritablement la complexité d'une compétence. En effet, le système de cotation des performances est, le plus souvent, analytique : le score à une tâche complexe n'est alors que la somme des scores attribués aux performances élémentaires.

2. LES DÉMARCHES DIAGNOSTIQUES CENTRÉES SUR LES PERFORMANCES

La majorité des tests d'évaluation diagnostique des apprentissages mathématiques se concentre sur les performances et souffre d'un manque cruel de modèle théorique permettant de comprendre les causes profondes des phénomènes observés. Comme le font justement remarquer Snow et Loh-

man (1989, p.268), *«les créateurs d'items sont le plus souvent des spécialistes du contenu qui travaillent sur base de spécifications de test sans aucune relation avec une théorie psychologique pertinente»*. Dans cette perspective, les items et les scores sont des boîtes noires qui n'offrent aucune explication des performances.

A titre d'illustration, nous pouvons prendre deux exemples de ce genre de test. Le premier est le *Wide Range Achievement Test -WRAT-* (Jastak & Wilkinson, 1984) qui est un test d'évaluation des apprentissages largement utilisé aux USA. Le manuel présente l'évaluation des troubles d'apprentissage comme une des fonctions de ce test. Selon l'auteur, *«il permet de déterminer précisément où un sujet a des difficultés et permet dès lors de prescrire des programmes de remédiation pour le traitement de ces faiblesses spécifiques»* (Jastak & Wilkinson, 1984, p.1).

Concrètement, pour les enfants de 5 ans à 11 ans 11 mois, la partie arithmétique du *WRAT* comprend 7 questions orales (essentiellement en rapport avec la numération) et 44 calculs écrits. Les scores à ces items permettent de calculer une note brute en arithmétique, qui peut ensuite être traduite sous forme de note standardisée, de percentile ou de niveau scolaire en référence aux résultats d'un échantillon représentatif de la population des élèves américains. Les différents items ont été sélectionnés sur base du modèle de Rasch. L'auteur postule donc un continuum latent unidimentionnel sur lequel les items peuvent être ordonnés et mis en correspondance avec un niveau d'habileté θ (thêta). Mais est-il défendable de considérer la compétence en arithmétique comme une réalité unidimentionnelle ? Par ailleurs, le caractère diagnostique de l'outil provient uniquement de la comparaison du score globale du sujet avec celui d'un groupe de référence. L'élève dont les résultats sont par trop inférieurs à ceux de ce groupe est considéré comme en difficulté, sans qu'aucune explication de cette situation ne puisse être avancée.

Le deuxième exemple est celui du *Key Math Revised* (Connolly, 1988). Cet instrument est explicitement présenté comme *«un inventaire diagnostique des bases en mathématiques»*. Par rapport au *WRAT*, le *Key Math* est beaucoup plus analytique. Trois grandes catégories de capacités sont évaluées : les concepts de base, les opérations et les applications. Chaque catégorie est elle même divisée en sous-catégories (par exemple, «addition», «soustraction»...) et en domaines (par exemple, «les faits arithmétiques», «les algorithmes pour additionner les nombres entiers», «l'addition des nombres rationnels»...). Ici aussi, le modèle de Rasch a été utilisé pour ordonner les items de chaque catégorie. Les résultats bruts sont transformés et comparés à ceux d'un échantillon représentatif de la population.

Bien que construit avec grand soin, le *Key Math* soulève des questions fondamentales. Le découpage en catégories et en sous-catégories n'est appuyé sur aucun modèle théorique. En fait, nous retrouvons les subdivisions traditionnelles présentées dans les programmes de mathématique de l'école primaire. Il s'agit d'un découpage de matière et non d'une analyse appuyée sur

un modèle cohérent de l'apprentissage de la mathématique. L'interprétation des résultats se fait non seulement au niveau des profils mais aussi au niveau des items dont l'auteur offre une description en terme d'objectifs comportementaux (par exemple, *«le sujet peut additionner des fractions de même dénominateur»*). Le point de vue reste ici très behavioriste. Aucune hypothèse n'est avancée à propos des processus utilisés par l'élève pour résoudre une problème. De même, aucune explication n'est proposée pour comprendre un échec. Ainsi, lorsque le sujet ne peut répondre correctement à la question *«2/5 + 1/5»*, nous pourrons simplement dire qu'il ne peut additionner les fractions de même dénominateur. Pourquoi ne peut-il y arriver ? Quelle relation peut-on faire entre cet échec et les autres performances du même élève ? Quels sont les pistes pour une éventuelle prise en charge remédiative ?... Nous n'en savons rien!

Comme nous pouvons nous en rendre compte au travers des deux exemples que nous venons brièvement de présenter, une approche en terme de performance se révèle rapidement insuffisante pour le diagnostic. Tout au plus permet-elle de réaliser une première appréciation de la situation. Mais elle n'offre pas de véritable compréhension des difficultés rencontrées par les sujets et, partant, n'offre guère d'assise solide à une prise en charge remédiative efficace. Celle-ci risque de n'être qu'une répétition des explications qui ont précédemment échoué.

3. LES DÉMARCHES DIAGNOSTIQUES CENTRÉES SUR LES COMPÉTENCES

3.1 L'approche piagétienne

3.1.1 *Le modèle piagétien du nombre et son application diagnostique*

Le modèle piagétien a occupé, et occupe encore, une place importante dans le diagnostic des troubles d'apprentissage en mathématique. Il permet en effet de dépasser le simple constat de réussite ou d'échec et de saisir le mode de raisonnement utilisé par le sujet. En ce sens, il offre une compréhension des phénomènes observés et un certain nombre de pistes pour les actions remédiatives. Malheureusement, nous allons voir que ce modèle séduisant n'est pas sans limite et qu'aujourd'hui, il ne peut tenir lieu d'unique référence pour le diagnostic.

Rappelons brièvement le point de vue piagétien sur le nombre. Comme le souligne Bideaud (1991, p.19), dans l'épistémologie piagétienne, *«le sujet est acteur, auteur même et constructeur de l'objet»*. Ce postulat de base détermine profondément le concept de nombre défendu par Piaget. Selon Piaget, le nombre n'est pas une réalité externe que l'enfant absorberait passivement. Au contraire, le nombre est construit par l'enfant grâce à ses capacités

logiques. Il est en effet évident que le nombre n'est pas une propriété des objets au même titre que la couleur ou la dureté. Le nombre est une relation entre les objets que crée le sujet par abstraction réfléchissante. Sa véritable source est interne au sujet.

Piaget propose ainsi une explication psychogénétique des opérations mathématiques où la logique joue le rôle d'un système normatif que construit progressivement l'enfant. Dans cette perspective, *«l'étude du développement du nombre sera celle des opérations «sur le nombre» et celle des sous-systèmes de sériation et de classification qui les fondent»* (Bideaud, 1991, p.20). Dans son ouvrage écrit en collaboration avec Szeminska, *«La genèse du nombre chez l'enfant»* (1941), Piaget s'attache à décrire ces capacités logiques que, progressivement, l'enfant acquiert et coordonne pour arriver à maîtriser le concept de nombre. Ces capacités se développent plus au moins parallèlement et leur intégration débute avant leur maîtrise complète. Décrivons-les brièvement.

L'opération de sériation consiste à ordonner les objets à partir de leurs différences. Il s'agit de distinguer les objets en ne tenant compte que d'une ou de plusieurs variables (la taille, le poids...). Le sujet crée donc une relation asymétrique entre les objets et fait abstraction de leurs équivalences. Au niveau numérique, la sériation est illustrée par l'acquisition de la suite ordonnée des nombres entiers naturels : 5 est plus grand que 4 qui lui-même est plus grand que 3. Lorsque l'enfant atteint le niveau opératoire de la sériation, il peut coordonner les relations inverse $n+1>n$ et $n< n+1$. Il peut donc dérouler la série dans les deux sens.

L'opération de classification consiste, elle, à ranger les objets en un ensemble commun en faisant abstraction de leurs différences et en ne prêtant attention qu'à leurs qualités communes. La relation créée est ici l'équivalence. Elle permet la construction de classes logiques. Au niveau numérique, la classification est à la base de l'aspect cardinal du nombre. Le nombre cardinal représente en effet un ensemble d'éléments considérés comme équivalents, en l'occurrence comme des unités. La maîtrise des opérations de classification évolue vers la capacité de réaliser des classifications hiérarchiques. Les classes peuvent en effet être emboîtées les unes dans les autres. L'enfant apprend dès lors à raisonner sur les relations entre les parties et le tout. Au niveau du nombre, cela se marque par la compréhension de l'inclusion des classes numériques : la classe «3» est incluse dans la classe «4» qui, elle-même, est incluse dans la classe «5».

Les différentes capacités logiques que nous venons de détailler vont progressivement se coordonner jusqu'à la maîtrise des nombres comme classes sériées. Car *«les nombre finis sont [...] nécessairement à la fois cardinaux et ordinaux et cela résulte de la nature même du nombre qui est d'être un système de classes et de relations asymétriques fusionnées en un même tout opératoire»* (Piaget et Szeminska, 1941, p.204). A ce stade, l'enfant comprend que les nombres sont des classes emboîtées mais égale-

ment ordonnées, qu'ils sont à la fois ressemblance (classe) et différence (ordre). Par exemple, les éléments qui composent le nombre 5 sont équivalents en tant qu'unités mais ils sont, en même temps, non équivalents car ils peuvent être ordonnés.

Et *la conservation* ? Jusqu'à présent, nous n'avons pas parlé de la conservation du nombre qui apparaît souvent comme un aspect essentiel à évaluer. En réalité, la conservation n'est pas une propriété du nombre qui viendrait s'ajouter aux propriétés cardinales et ordinales. La conservation doit plutôt être vue comme constitutive des caractéristiques cardinales et ordinales du nombre. Rappelons que pour qu'un enfant soit conservant, *«la connaissance empirique doit être interprétée et corrigée par la raison»* (Kamii, 1990, p.38). Pour pouvoir se dégager des apparences sensibles, la pensée doit être réversible : elle doit être capable d'exécuter la même opération dans les deux sens de parcours, en ayant conscience qu'il s'agit de la même action. Ainsi, l'enfant comprend que la modification de la disposition spatiale des éléments ne modifie nullement le cardinal d'un ensemble dénombré. Il peut en effet revenir mentalement à la situation initiale et se dégager ainsi des changements de surface. Si nous ne retirons ni n'ajoutons rien à un ensemble, celui-ci conserve le même cardinal malgré les apparences.

Bien que créé voici 50 ans, le modèle piagétien du nombre reste toujours une référence pour l'évaluation diagnostique. Certaines erreurs des élèves ne peuvent être réellement comprises qu'en lien avec la maîtrise du concept de nombre. Trop souvent, l'attention des enseignants se focalise sur les procédures au détriment de la notion du nombre que possèdent les enfants. Or des erreurs fréquentes découlent avant tout d'une mauvaise compréhension du nombre sur lequel sont effectuées les opérations arithmétiques (Jonnaert, 1994, pp. 225-233).

Prenons l'exemple d'une soustraction écrite dont le résultat est erroné :

$$\begin{array}{r} 506 \\ \underline{-\ 42} \\ 544 \end{array}$$

Si nous analysons l'erreur en terme de procédure, nous pouvons avancer l'hypothèse qu'estimant ne pas pouvoir effectuer *«0 - 4»*, le sujet a inversé les termes et effectué *«4 - 0»*. Le sujet n'aurait donc pas appliqué la procédure correcte qui consiste à réaliser un emprunt dans le rang des centaines pour le rang des dizaines. Or, l'élève qui effectue une différence entre deux entiers positifs doit se rendre compte que le résultat recherché doit *nécessairement* être inclus dans le premier terme de l'opération. Dans l'exemple ci-dessus, le résultat doit logiquement être inclus dans *«506»*. Par conséquent, lorsque l'élève propose *«544»* comme réponse, il ne commet pas seulement une erreur de procédure mais aussi une faute de logique. Sans effectuer l'opération, il devrait reconnaître que cette réponse ne peut pas être correcte.

Ce bref exemple, nous démontre l'intérêt de la référence au modèle piagétien dans le cadre du diagnostic des difficultés en mathématique. Mais, pratiquement, comment conduire une évaluation diagnostique fondée sur le modèle piagétien ? Une première possibilité est de se référer aux ouvrages de Piaget et d'en extraire les épreuves qui apparaissent les plus adéquates. Une autre possibilité est d'utiliser une batterie d'épreuves construites sur base du modèle piagétien, comme l'UDN 80 (Meljac, 1980). Cette batterie est présentée comme *«un échantillonnage [...] d'occasions au cours desquelles l'enfant examiné fait preuve de ses possibilités de maniement du nombre et des opérations mentales qui y sont associées»* (Meljac, 1980, p. 9). Son objectif est explicitement d'aider à comprendre les difficultés d'apprentissage en mathématique et de servir de guide lors des prises en charge rééducatives. L'UND 80 est composée de 13 épreuves regroupées en quatre catégories : (1) des épreuves concernant les structures logiques élémentaires (classification et sériation); (2) des épreuves de conservation; (3) des épreuves portant sur l'utilisation des nombres dans des situations variées et (4) des épreuves mettant en cause la notion «d'origine», de point de départ d'une action. Dans la démarche diagnostique, la mise en relation des résultats aux différentes épreuves joue un rôle essentiel. Les rapports existant entre les performances aux épreuves de type logique et l'utilisation des premières notions mathématiques doit faire l'objet d'une attention particulière.

3.1.2 *Les limites du modèle piagétien*

·Importer la théorie opératoire dans le domaine de l'évaluation est malheureusement moins simple qu'il n'y paraît au premier abord. Comme le rappelle, de façon quelque peu provoquante, Claire Meljac (1988, p.200) : *«Contrairement aux idées reçues, Piaget n'a jamais été un psychologue- encore moins un psychologue de l'enfance- moins encore un psychologue de l'enfant»*. Et il est vrai que, dès l'origine et tout au long de sa vie de chercheur, le point de vue de Piaget n'a pas cessé d'être, avant tout, celui d'un philosophe. Plus précisément, celui d'un philosophe de la connaissance. Son but est en effet *«de reconstituer la genèse ou les phases de formation de l'intelligence, jusqu'à pouvoir rendre compte du niveau opératoire final»* (Piaget, 1947, p.55). Ses recherches sur les enfants n'ont été qu'un moyen pour atteindre cet objectif. Les épreuves qu'il a imaginées avec ses collaborateurs n'ont donc jamais eu pour but le diagnostic des troubles d'apprentissage. L'application de ses épreuves en psychologie différentielle et en éducation ne faisait pas partie de ses intérêts, bien qu'il ait généralement regardé avec bienveillance les tentatives en ce sens réalisées par d'autres[2]. Extraire des épreu-

2 *«L'idée d'appliquer au diagnostic et au pronostic des arriérations, de la débilité mentale et des troubles psychopathologiques infantiles en général, les résultats des études sur le développement des structures opératoires chez l'enfant normal s'est révélée féconde et même indispensable dans la mesure où ces études atteignent le dynamisme formateur de l'intelligence au sens le plus large»* (Piaget, préface à la seconde édition de l'ouvrage d'Inhelder, «Le diagnostic du raisonnement chez les débiles mentaux», 1943, p.2).

ves d'un de ses ouvrages et les appliquer sans autre forme de procès, est donc un peu rapide. Ces épreuves n'ont pas été créées dans une optique différentielle et ne possèdent pas, dans l'état où nous les lègue Piaget, toutes les qualités nécessaires pour occuper la fonction d'instrument d'évaluation. Diverses précautions méthodologiques s'imposent (cfr Grégoire, 1992, pour une analyse détaillée de cette question).

Certaines concernent la standardisation des procédures. Evaluer est toujours comparer, que ce soit à un objectif d'apprentissage ou aux performances d'autres élèves. Si les conditions dans lesquelles les sujets passent les épreuves ne sont pas identiques, des comparaisons valides ne sont plus possibles.

Un autre problème méthodologique concerne la définition des conduites intermédiaires. Dans une perspective diagnostique, il ne suffit souvent pas de classer les conduites selon qu'elles sont de niveau opératoire au non. Une appréhension plus fine du niveau de développement cognitif est souvent utile. Les étapes intermédiaires décrites par Piaget peuvent être des points de repère intéressants. Mais ils ne sont pas toujours suffisants ni même pertinents. Par exemple, les travaux de Tollefsrud-Anderson et al. (1991) sur la conservation du nombre mettent clairement en évidence une étape du développement où la réussite de l'épreuve de conservation est acquise grâce au comptage (et non grâce au seul raisonnement logique). Lors de cette étape, le comptage est utilisé comme une technique de preuve qui permet à l'enfant de conforter la validité de son raisonnement. Lorsque la solidité de ce raisonnement est suffisamment assurée, l'enfant peut alors abandonner la vérification par le comptage.

Cet exemple renvoie à la question la plus épineuse posée par l'évaluation diagnostique basée sur le modèle piagétien : celui de la validité. Les épreuves piagétienne mesurent-elles bien ce qu'elles prétendent mesurer, et rien que cela ? Cette question rejoint celle que nous avons déjà discutée dans le § 1. Dès le moment où notre évaluation se concentre sur les compétences, et non sur les performances, nous devons être certains que les tâches choisies sont de bons révélateurs de ces compétences. Or, de nombreux travaux sur les épreuves piagétiennes indiquent que cette exigence est difficile à rencontrer. Ainsi, Fayol (1990), dans une importante recension des recherches sur l'évaluation de la conservation du nombre, remarque que les résultats mettent en évidence l'impact, parfois dominant, de variables considérées par Piaget comme secondaires. C'est par exemple le cas de la formulation verbale des questions et de la disposition visuo-spatiale des ensembles présentés.

Un constat similaire est fait par Bideaud (1988) à propos de l'opération de classification. Classiquement, la maîtrise opératoire de la classification est évaluée au travers de la quantification de la relation d'inclusion par la question : *«Y a-t-il plus de A que de B ?»*, B étant un sous-ensemble inclus dans l'ensemble A. L'enfant de niveau opératoire arrive à la solution correcte grâce à un raisonnement du type : $A = B + B'$ et $B = A - B'$, donc $A > B$. A priori,

nous devrions obtenir des résultats similaires à toutes les épreuves faisant appel au même raisonnement. Or, ce n'est pas ce qui est observé. Les pourcentages de réussite rapportés par Piaget lui-même varient sensiblement selon les épreuves. Par la suite, de nombreux chercheurs ont repris les épreuves d'inclusion soit dans leur forme originale soit légèrement modifiées. Leurs résultats confirment que des facteurs parfois minimes peuvent rendre la tâche plus ou moins difficile et affecter notablement les réponses des enfants (la nature perceptive ou verbale du matériel utilisé, l'extension des sous-classes, la prégnance perceptive de la classe emboîtante,...). Partant de ce constat, Bideaud (1988, pp.97-98) conclut qu' *«il est difficile de considérer la réponse correcte et justifiée d'inclusion comme le critère de la classification opératoire, dans sa relation avec la sériation et le nombre»*. Et elle ajoute *«[le] concept même de «décalage horizontal» qui est invoqué implicitement à partir du moment où une bonne réponse d'inclusion n'en suscite pas nécessairement une autre, paraît une carte forcée pour légitimer a posteriori un état de fait»*.

Est-il dès lors impossible d'évaluer de manière valide les capacités logiques des sujets ? Nous ne pouvons en effet évaluer un raisonnement pur. Celui-ci s'applique toujours à un contexte particulier qui influence positivement ou négativement les performances des sujets. Comme nous ne pouvons éliminer cette influence du contexte, la seule solution est de suivre la proposition de Messick et de *«transcontextualiser»* notre évaluation. Seule la consistance des performances au travers de différents contextes peut être prise comme un indicateur valide du niveau de compétence des élèves. Ce point de vue sur la démarche évaluative est d'ailleurs celui qui est défendu par Meljac (1980, p. 53) qui souligne l'importance pour le diagnostic de *«la cohérence interne des réussites de l'enfant»*.

La question du contexte renvoie à une limite fondamentale du modèle piagétien. Comme nous l'avons déjà souligné, Piaget n'est pas psychologue mais avant tout philosophe de la connaissance et logicien. Son point de vue sur le développement cognitif est normatif. Il essaie en effet, tout au long de son oeuvre, de définir les normes logiques nécessaires à la construction et à l'évolution de la connaissance. Mais en procédant de la sorte, Piaget néglige les conditions socio-culturelles et spécifiquement psychologiques qui influencent inévitablement la construction du nombre. Il le reconnaît d'ailleurs clairement : *«Il va de soit qu'avant de chercher à traduire la chose sous une forme statistique et d'appliquer à ces épreuves les diverses formules de corrélations du calcul des probabilités, il faudrait résoudre auparavant des questions pour lesquelles nous avouons n'avoir aucun intérêt : tel problème d'ordination dont nous nous sommes servis est-il exactement de difficulté égale à tel autre ou à tel problème de cardination, indépendamment de l'ordination et de la cardination elle-même ? Il est clair en effet, qu'en chaque épreuve intervient une foule de facteurs hétérogènes, tels les mots employés, la longueur de la consigne, son caractère plus ou moins concret, ses relations avec l'expérience individuelle*

du sujet, le nombre des objets considérés, l'intervention de la numération apprise, etc., etc.» (Piaget & Szeminska, 1941, p.193).

Si, pour le logicien, l'aspect pragmatique de la connaissance est sans doute peu important, il n'est est pas de même pour le psychologue et l'enseignant. Le but de ces derniers n'est pas de construire un modèle général du sujet épistémique mais d'agir avec un sujet concret dans un contexte particulier. Cette différence de niveau d'approche de la construction de la pensée permet de comprendre le relatif échec de l'application des épreuves piagétiennes dans les situations concrètes d'évaluation.

La solution n'est toutefois pas de rejeter le modèle piagétien du nombre dont nous avons pu souligner la valeur mais, plutôt, de le réformer. Il s'agit en effet d'y introduire le sujet réel et son contexte et d'élaborer ce que Bideaud & Houdé (1991, p.90) appellent *«un constructivisme-en-contextes»* où seraient articulés le contexte pragmatique et le contexte normatif. Les développements récents de la psychologie cognitive nous offrent les outils conceptuels pour nous permettre de réaliser cet élargissement du modèle constructiviste classique.

3.2 Les approches cognitivistes

3.2.1 *Apports des modèles cognitivistes*

Comme le souligne Richelle (1993, p.93), *«il n'y a pas un cognitivisme, il y a des variantes de cognitivismes»*. Sont regroupés sous un même terme des modèles théoriques parfois très différents mais qui se rejoignent par une volonté commune de dépasser le behaviorisme en étudiant le traitement de l'information qui est réalisé entre le stimulus et la réponse. Assez curieusement, la plupart des auteurs cognitivistes qui travaillent sur les apprentissages mathématiques ne font aucune référence à Piaget. Pourtant le point de vue piagétien, s'il est d'évidence moins «psychologique», n'est pas en contradiction avec celui des cognitivistes. Il est, au contraire, une source d'informations essentielle pour nous aider à comprendre certaines erreurs et certaines difficultés d'apprentissage en mathématique. Comme nous le verrons plus loin cette ignorance des travaux piagétiens conduit certains chercheurs à redécouvrir ce qui avait déjà été mis en évidence cinquante ans plus tôt par Piaget et ses collaborateurs.

Par rapport au point de vue piagétien, les approches cognitivistes sont plus soucieuses de comprendre le sujet réel en contexte. Ainsi, lorsqu'ils étudient la résolution de problèmes exprimés verbalement, les chercheurs veulent comprendre le traitement par le sujet de l'information dans toute sa complexité, sans se limiter à saisir les seules composantes opératoires. Ils tiennent dès lors compte de la coordination par le sujet de différentes connaissances : mathématiques, linguistiques, factuelles... Les chercheurs sont également plus attentifs aux représentations et aux types de connaissances utilisés par les sujets. En particulier, la distinction entre les connaissances

déclaratives et procédurales ainsi que les transformations de l'une en l'autre ont été largement appliquées pour décrire l'évolution des connaissances mathématiques. Ceci va de pair avec une grande attention portée au degré d'automatisation des procédures; laquelle est évaluée en terme de vitesse de traitement et/ou de résistance aux interférences d'autres tâches réalisées simultanément. Cette attention rejoint celle, plus large, portée aux contraintes du système cognitif : la capacité limitée de la mémoire de travail, la nécessité de pouvoir ignorer les informations non pertinentes et d'inhiber les procédures incorrectes...

3.2.2 *Modèles en composantes et évaluation*

Plusieurs modèles cognitivistes proposent des représentations du fonctionnement cognitif sous forme d'une organisation de composantes, chacune responsable d'un traitement spécifique de l'information. Ces modèles en composantes constituent des outils conceptuels importants pour l'évaluation diagnostique. Ils représentent un guide pour la démarche évaluative et une référence pour l'interprétation des performances observées. Dans la présente section, nous présenterons deux de ces modèles. Le premier concerne la résolution de problèmes. Le second, développé dans le cadre de la neuropsychologie cognitive, est plus focalisé et ne concerne que l'arithmétique élémentaire.

A. LA RÉSOLUTION DE PROBLÈMES

Mayer (1985) distingue quatre étapes dans le traitement d'un problème mathématique :

1. La traduction du problème

Chaque proposition du problème est traduite en une représentation interne. Cette traduction suppose des connaissances linguistiques et des connaissances factuelles (par exemple, qu'une livre de beurre vaut 500 grammes). On constate que les jeunes enfants ont souvent de difficultés à se représenter des propositions relationnelles (Riley & al., 1982); par exemple : *«Jean a 3 billes, Paul a 5 billes de plus que Jean. Combien Paul a-t-il de billes ?»*. Souvent les enfants se font une représentation statique de la situation (Paul à 5 billes) et non dynamique (billes de Paul = 5 + billes de Jean).

2. L'intégration du problème

Les propositions du problème sont mises ensemble au sein d'une représentation cohérente. Ceci suppose la connaissance de schémas, c'est-à-dire de formes typiques de problèmes. Partant de l'énoncé linguistique, cette intégration n'est pas toujours simple. Par exemple, les trois énoncés suivants doivent conduire à la même opération arithmétique (3+5=...) mais le premier se révèle beaucoup plus simple à intégrer que les deux suivants : (a) *«Jean a 3 billes. Paul lui en donne 5 de plus. Combien de billes Jean a-t-il à présent ?»* ; (b) *«Jean a 3 billes. Paul a 5 billes. Combien de billes ont-ils*

ensemble ?»; (c) *«Jean a 3 billes. Paul a 5 billes de plus que Jean. Combien Paul a-t-il de billes ?».*

3. La planification de la solution

Le sujet élabore une stratégie de résolution du problème qu'il a traduit en un énoncé mathématique. Par exemple, lors de l'apprentissage de l'algèbre, les élèves acquièrent une connaissance stratégique relative à la hiérarchie des opérations à effectuer : d'abord les parenthèses, puis les puissances, ensuite les multiplications et divisions et, enfin, les additions et soustractions.

4. L'exécution de la solution

Les calculs sont réalisés selon la stratégie choisie. Par exemple (Groen & Parkman, 1972), on peut résoudre une addition en : (a) recomptant tout; (b) comptant à partir du premier nombre de l'énoncé; (c) comptant à partir du plus grand nombre. Chez les sujets plus âgés, le calcul peut être évité par le rappel de «faits arithmétiques», c'est-à-dire de résultats d'opérations fréquentes mémorisés au cours du temps.

Un tel modèle est intéressant pour le diagnostic car il permet de mieux situer les difficultés du sujets sans se limiter à l'aspect strictement mathématique de la résolution d'un problème. Ce modèle a par ailleurs le mérite d'attirer l'attention sur deux facettes importantes de la résolution de problèmes : (1) le passage du «nombre de» au «nombre» et (2) l'importance des caractéristiques mnésiques du sujet.

Dans la vie quotidienne, nous ne rencontrons que des «nombres de», c'est -à-dire des quantités (30 litres d'essences, 600 grammes de viandes, 8 pommes, 245 francs...). Ces quantités, qui peuvent être discrètes ou continues, ne peuvent être traitées directement au niveau arithmétique. Pour ce faire, nous devons faire abstraction des qualités sensibles des objets et ne plus conserver que leur caractère d'unité. Comme le fait remarquer Popper (cité par Jonnaert, 1994, p.253), au niveau physique, deux gouttes plus deux gouttes ne font pas quatre gouttes dans une bouteille. Et un lapin plus un lapin peut donner plus de deux lapins. Par contre, au niveau mathématique, 2 + 2 fait toujours 4 en vertu des définitions de «2», «4», «+» et «=». En fait, l'énoncé 2+2=4 est une tautologie puisque nous affirmons simplement qu'une certaine manière d'écrire les choses est équivalente à une autre. Le passage du monde physique au monde mathématique est souvent source de difficultés pour les enfants qui se dégagent avec peine des apparences sensibles. Par ailleurs, les nombres étant par nature abstrait, se produit une rupture dans le traitement des problèmes. Il y a là danger de perte de sens. Le sujet entre dans un monde qui lui paraît mystérieux, voire magique. Il applique des algorithmes sans trop les comprendre et produit une solution absurde au regard de la réalité.

Le modèle présenté attire également l'attention sur les contraintes de fonctionnement du système cognitif. La capacité de la mémoire de travail est

en effet limitée. Comme il est nécessaire de traiter séquentiellement les informations, la vitesse de ce traitement doit être suffisante si l'on veut éviter d'oublier les résultats intermédiaires en mémoire de travail. Ces contraintes n'ont été guère prises en compte par Piaget. Pourtant, dans la réalité du calcul, on constate que les sujets qui n'ont pas suffisamment stocké de faits arithmétiques en mémoire à long terme surchargent leur mémoire de travail par de nombreuses opérations intermédiaires. La charge mentale de la résolution d'un problème arithmétique est évidemment moindre si le sujet a déjà en mémoire le résultat des additions et des multiplications les plus fréquentes.

B. L'ARITHMÉTIQUE

Les travaux réalisés dans le domaine de la neuropsychologie cognitive sont également une source intéressante de connaissances à propos des composantes qui interviennent dans le traitement de problèmes mathématiques. En neuropsychologie cognitive, les patients atteints d'une lésion cérébrale sont analysés moins dans le but de mettre à jour l'origine neurologique d'un dysfonctionnement mental que de comprendre la cognition humaine normale. Dans cette perpective, un trouble est conçu comme *«le résultat visible d'un système de traitement de l'information au sein duquel certaines composantes du traitement sont soit devenues dysfonctionnelles soit se trouvent déconnectées des autres composantes»* (Seron, 1991). Cette conception des choses s'appuie sur le postulat d'une autonomie fonctionnelle des différentes composantes et donne lieu à une démarche d'évaluation diagnostique qui consiste à tester systématiquement le fonctionnement des différentes composantes du modèle.

Bien que développés à partir d'étude de cas d'adultes souffrant de déficits acquis des fonctions cognitives, les modèles neuropsychologiques peuvent être également utiles pour comprendre les troubles d'apprentissage. Ces modèles offrent en effet une représentation de l'organisation du fonctionnement cognitif normal, laquelle se construit progressivement au cours des apprentissages. Elle peut dès lors constituer une référence pour l'évaluation diagnostique des enfants.

Dans le domaine mathématique, plusieurs modèles ont été proposés. Actuellement, le modèle à la fois le plus élaboré et le plus opérationnel est dans doute celui de Mc Closkey (1992). Ce modèle (figure 1) se présente sous la forme d'une architecture de trois systèmes : (1) un système de compréhension des nombres, (2) un système de calcul et (3) un système de production des nombres. Les systèmes de compréhension et de production sont caractérisés par une structure semblable. Il contiennent tous les deux un sous-système verbal et un sous-système de nombres arabes. Une même représentation sémantique de la valeur du nombre est produite quel que soit le code d'entrée ou de sortie. Le système de calcul est lui divisé en (1) un sous-système d'interprétation des symboles qui précisent l'opération à effectuer; (2) un sous-système de recherche de faits arithmétiques (par exemple, les tables de multiplication) et (3) un sous-système d'exécution des calculs activé lorsque la

réponse ne se trouve pas dans le stock de faits arithmétiques. Ces différents sous-systèmes permettent de distinguer, au sein des connaissances arithmétiques, les connaissances déclaratives et les connaissances procédurales.

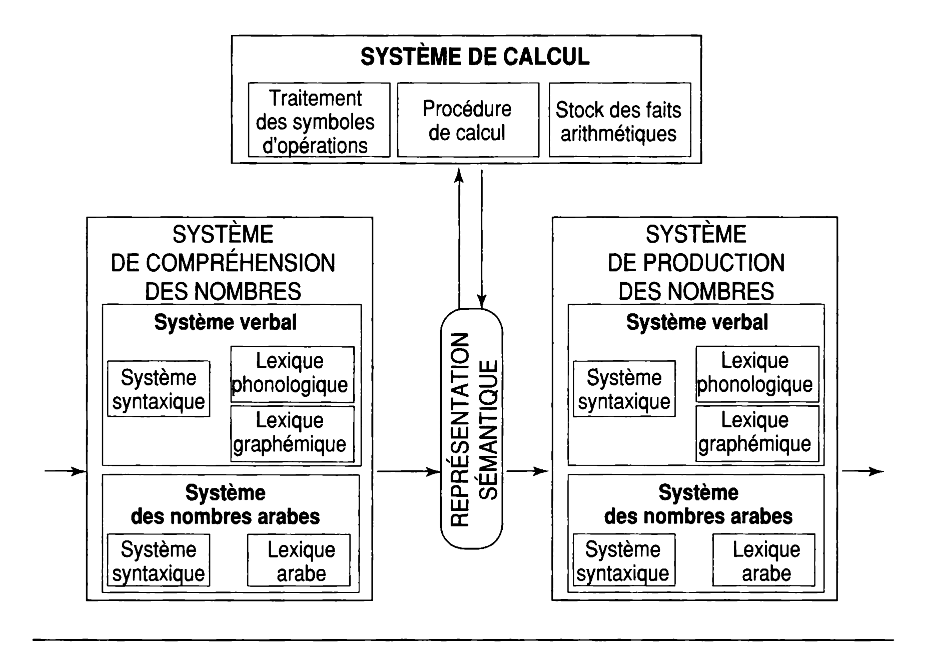


FIGURE 1

Modèle du traitement des nombres et du calcul selon McCloskey

Nous pouvons illustrer le fonctionnement de cette organisation modulaire par l'exemple du traitement d'un calcul arithmétique. Imaginons un élève qui doit résoudre le calcul *«5+6»* qu'il vient de lire au tableau. Les étapes du traitement seront les suivantes :

1. dans le système de compréhension, cinq et six sont transformés en des quantités abstraites par le composant lexical du sous-système des nombres arabes.
2. dans le système de calcul, le sous-système d'interprétation des symboles identifie l'opérateur de l'addition. Ensuite, le résultat est recherché dans le stock des faits arithmétiques.
3. dans le système de production, la représentation du résultat est transformée dans la forme phonologique adéquate. L'élève peut alors émettre sa réponse oralement.

Comme le souligne Seron (1994, p. 443) : *«La valeur heuristique du modèle est évidente car il permet d'analyser de manière relativement précise quelle composante se trouve déficitaire»*. Sur cette base, il est possible de créer des tâches spécifiques afin de déterminer quelles sont les com-

posantes efficientes et celles qui sont déficitaires. L'interprétation des résultats ne se fait pas épreuve par épreuve mais sur base des prédictions coordonnées qui peuvent être faites à l'aide du modèle.

3.2.3 *Limites des démarches diagnostiques cognitivistes*

Les travaux réalisés en psychologie cognitive ont généré des modèles intéressants pour l'évaluation diagnostique. Toutefois, en se focalisant sur les procédures, ces modèles négligent souvent l'objet sur lequel portent les opérations mathématiques : le nombre. Or, c'est la compréhension de la structure profonde de ce dernier qui permet au sujet de choisir la stratégie de résolution la plus adéquate et de contrôler la valeur logique de la solution trouvée. Par exemple, lorsqu'il opère sur les entiers naturels, l'élève qui soustrait un nombre d'un autre ne peut trouver comme solution une différence qui soit plus grande que les termes de celle-ci. De même, l'élève qui résout l'addition *«3+5=... »* en comptant à partir du plus grand nombre démontre ainsi sa compréhension de la loi de commutativité de l'addition, selon laquelle A+B = B+A. Or la compréhension de cette loi «exige *à la fois toute la structure de la pensée opératoire, et toute la structure du nombre proprement dit»* (Gréco, 1962, p.226).

Peu attentifs à la construction des structures logiques sous-jacentes à la résolution des problèmes mathématiques, certains chercheurs se perdent ainsi dans des explications qui ne touchent qu'à la périphérie de la question. Par exemple, Stern (1993) tente de comprendre pourquoi le problème *«Jean a 7 oeufs. Il a 4 oeufs de plus que Pierre. Combien d'oeufs Pierre a-t-il ?»* est beaucoup plus difficile pour les enfants que le problème *«Jean a 7 oeufs. Pierre en a 4 de moins que Jean. Combien d'oeufs Pierre a-t-il ?»*. Différentes expériences sont conduites afin de déterminer la source de cette difficulté. On observe que pour résoudre le problème, le sujet peut simplement transformer le premier énoncé dans la forme du second. L'auteur explique que cette transformation requière *«une flexibilité dans l'usage du langage décrivant les quantités»* (Stern, 1993, p. 21). Elle reconnaît toutefois que le manque de flexibilité du langage peut être dû au fait que, pour le jeune enfant, l'addition et le soustraction sont compris comme des processus indépendants et non comme des opérations réciproques. Et elle ajoute :*«Les relations entre addition et soustraction ne peuvent être elles-mêmes comprises qu'au moment où les nombres sont représentés en termes de relations entre les parties et le tout»* (p.21). Cette observation est pour le moins surprenante. En effet, l'auteur, qui semble ignorer totalement les travaux de Piaget, est conduite à redécouvrir de manière embryonnaire les fondements opératoires du nombre et des opérations décrites à Genève voici plusieurs décennies. Ce bref exemple illustre combien il est difficile de faire l'économie de l'apport piagétien pour la compréhension et l'évaluation des premiers apprentissages mathématiques. La résolution de problèmes mathématiques n'est pas uniquement affaire de procédure, elle est également guidée par les exigences d'un système opératoire qui se construit petit à petit au cours de l'enfance.

Par ailleurs, les modèles développés actuellement en psychologie cognitive sont généralement fort restreints, ne concernant que la résolution de certaines tâches bien précises. Chaque modèle local est utilisé pour étudier un problème spécifique à l'aide d'épreuves créées pour cette unique fonction. Ces modèles locaux constituent des *«minithéories»* (Richelle, 1993, p.221) qui ne sont généralement pas articulées entre elles. Or nous ne pouvons exclure que certains troubles d'apprentissage soient dus à des problèmes transversaux dont le retentissement concerne le fonctionnement cognitif de manière globale. Ces problèmes transversaux peuvent, par exemple, concerner l'attention ou de le fonctionnement mnésique. Or les modèles en composantes ont quelques difficultés à intégrer des variables d'ordre plus général. Il y a dès lors un risque de réductionnisme à vouloir s'enfermer dans un modèle local sensé rendre compte de toutes les facettes des compétences mathématiques.

Conclusion

Après ce passage en revue des différentes approches du diagnostique des apprentissages mathématiques, quelle démarche d'évaluation diagnostique pouvons nous retenir ? Nous avons pu constater que les démarches centrées sur les compétences, qu'elles soient d'inspiration piagétienne ou cognitiviste, nous apportent des informations plus riches et plus intéressantes pour les actions rééducatives que les démarches d'évaluation centrées sur les performances. Malheureusement, les premières sont nettement plus complexes et plus coûteuses en temps que les secondes. C'est pourquoi, nous pensons que les évaluations normatives centrées sur les performances gardent une place dans l'examen diagnostique. Elles permettent en effet de réaliser un passage en revue rapide de la situation de l'élève. En cas de problème, une investigation plus approfondie peut alors être conduite dans un secteur précis des apprentissages.

Par ailleurs, le développement de modèles cognitivistes n'implique pas que le modèle piagétien soit aujourd'hui dépassé. Chaque modèle apporte des informations spécifiques pour permettre de comprendre les caractéristiques du fonctionnement mental sous-jacent aux performances observées. Le modèle piagétien nous offre un cadre conceptuel particulièrement solide pour comprendre la construction du nombre par l'enfant. Tout d'abord, il attire notre attention sur la maîtrise d'opérations qui constituent les racines du concept de nombre. Ensuite, il met l'accent, au travers de la notion de groupement, sur la coordination progressive des opérations. Ce groupement des opérations est dirigé par la nécessité interne d'équilibre, c'est-à-dire de cohérence. Les différentes opérations forment ainsi un système dont découle la rigueur logique de la pensée. L'existence d'un tel groupement d'opérations nous permet de comprendre, entre autres, les relations logiques qui existent entre le concept de nombre et les différentes opérations arithmétiques. C'est en ce sens que l'on peut considérer le modèle piagétien du nombre comme un système normatif.

Mais l'élève confronté à des problèmes mathématiques n'est pas seulement un petit logicien. Il est aussi un sujet en contexte. Par conséquent, l'évaluation des capacités cognitives liées aux apprentissages mathématiques ne peut se limiter aux capacités opératoires. Les cognitivistes ont démontré la nécessité de prendre en compte d'autres caractéristiques de la pensée, en particulier la manière dont les connaissances sont représentées en mémoire. Par ailleurs, ils attirent l'attention sur les contraintes de fonctionnement du système cognitif. Par exemple, le fait que la mémoire de travail ait une capacité limitée nous conduit à évaluer le degré d'automatisation de certaines connaissances. Ces apports de la psychologie cognitive n'entrent pas en conflit avec une évaluation piagétienne des compétences mathématiques. Ils viennent, au contraire, utilement la compléter.



Niveaux préalables et apprentissages numériques élémentaires

Jacqueline Bideaud

Introduction

L'évaluation d'un apprentissage requiert la prise en compte de trois facteurs : le niveau préalable des enfants, le contenu même de l'apprentissage du point de vue des connaissances et des procédures *réellement* engagées et, enfin, le critère d'acquisition retenu (effet à cours ou à long terme, généralisation à des situations analogues etc...). L'estimation du niveau préalable est à l'évidence nécessaire, et cela en amont comme en aval. En amont elle permet d'orienter la mise en oeuvre de situations d'entraînement pertinentes. En aval elle assure une estimation plus exacte des effets d'apprentissage et une compréhension plus fine des erreurs. L'objectif de cet exposé est de montrer l'importance d'une évaluation la plus exhaustive possible des capacités «mathématiques» des jeunes enfants dans leur appropriation de la chaîne numérique verbale, la maîtrise des propriétés de cette séquence étant la condition *sine qua non* d'un *bon* apprentissage de l'arithmétique élémentaire.

La diversité des données expérimentales qui concernent la construction et l'utilisation du nombre chez l'enfant est extrême. Elle rend compte de la diversité des conquêtes cognitives qui jalonnent le développement dans ce domaine. C'est en ce sens que l'on peut parler d'une psychogenèse polymorphe ou encore avec Droz (1991 p. 299) «...des psychogenèses des notions relatives aux nombres naturels». Cette complexité, qui semble bien être la règle, ne facilite guère l'estimation des compétences numériques à un moment donné du développement. La difficulté est d'autant plus grande que la nature des conduites observées est diversement interprétée selon les points de vue théoriques et le rôle attribué aux contextes extra et/ou intra-scolaires d'acquisition.

Dans la première partie de cette contribution, j'essaierai de montrer comment un regard à la fois critique et synthétique sur les traits essentiels des théories les plus marquantes permet, compte tenu des faits expérimentaux disponibles, de guider plus sûrement les investigations. La seconde partie sera consacrée à l'exposé de trois études empiriques des capacités numériques d'enfants de 4 à 6 ans, retenues parce qu'elles sont les seules d'importance certaine et parce que l'analyse de leurs résultats conduit à des hypothèses constructives. Dans la troisième partie je tenterai de montrer l'importance des contextes dans l'estimation des niveaux préalables et, partant, dans celle des effets d'apprentissages.

1. UN REGARD SUR LES THÉORIES

1.1 La théorie piagétienne

Selon Piaget, la maîtrise du nombre n'est atteinte que lorsque l'enfant synthétise en un seul *groupe* les opérations de classification et de sériation. Il

est facile et courant de «malmener» la pensée piagétienne : aussi pour plus de précision citons-nous Piaget lui-même (1947, pp.169-170). :

«*Sans doute le jeune enfant n'attend pas cette génération opératoire pour construire les premiers nombres... mais les nombres de 1 à 6 sont encore intuitifs, parce que liés à des configurations perceptives. D'autre part on pourra apprendre à l'enfant à compter, mais l'expérience nous a montré que l'usage verbal des noms de nombre reste sans grand rapport avec les opérations numériques elles-mêmes, celles-ci précédant parfois la numération parlée ou lui succédant sans lien nécessaire. Quant aux opérations constitutives du nombre, c'est à dire à la correspondance bi-univoque (avec conservation de l'équivalence obtenue, malgré les transformations de la figure) ou a l'itération simple de l'unité (1+1=2; 2+1=3 etc...), elles ne requièrent pas autre chose que les groupements additifs de l'emboîtement des classes et de la sériation des relations asymétriques (ordre), mais fondus en un seul tout opératoire, tel que l'unité 1 soit simultanément élément de classe (1 compris dans 2; 2 en 3 etc...) et de série (le premier 1 avant le deuxième 1 etc...). Tant que le sujet envisage les éléments individuels dans leur diversité qualitative il peut, en effet, ou les réunir selon leurs qualités équivalentes (il construit alors des relations de classes) ou les ordonner selon leurs différences (il construit alors des relations asymétriques) mais il ne peut les grouper simultanément en tant qu'équivalents et que différents. Le nombre est au contraire une collection d'objets conçus comme à la fois équivalents et sériables, leurs différences se réduisant alors à leur position d'ordre : cette réunion de la différence et de l'équivalence suppose en ce cas l'élimination des qualités, d'où précisément la constitution de l'unité homogène 1 et le passage du logique au mathématiques*».

La citation est longue, mais tout y est : la logique sous-jacente, l'abstraction nécessaire et le bien fondé de l'épreuve de conservation numérique.

Il est évident que la perspective est incomplète. L'enfant baigne dans un environnement chiffré et les pratiques sociales guident et orientent sa conquête du nombre. Il est évident aussi que l'enfant quantifie avant de compter et que la théorie sous estime ce tremplin (Vilette, 1993, 1994). Mais il est non moins évident que la conservation de la quantité, en dépit de transformations qui la laisse invariante, reste un des critères sûrs de la compétence numérique. On rappelle que, dans l'épreuve piagétienne de conservation, 6 à 8 jetons sont alignés devant l'enfant à qui l'on demande de placer sur la table autant de jetons pris dans un tas d'une autre couleur. Vers 4 ans, l'enfant construit généralement, en serrant plus ou moins les jetons, une rangée de même longueur que celle de la rangée-modèle. Plus tard, l'enfant met en correspondance terme à terme les jetons des deux rangées. Mais si l'adulte éloigne le dernier jeton de la rangée initiale, l'enfant dénie l'équivalence et rajoute un ou plusieurs jetons à la sienne. Vers 5-6 ans, il convient de l'équivalence des quantités quelles que soient les transformations apparentes opérées. Il est alors «conservant». Gréco (1962), bien avant Gelman (1982),

Gelman et Meck (1983, 1991), observe, lors d'une épreuve où, chacun d'un côté de l'écran, l'adulte et l'enfant mettent en correspondance cinq ou six jetons comptés, que l'invariance numérique est maintenue précocement («cela fait 6 et 6», par exemple). Mais l'écran enlevé, l'invariance de la quantité est cependant réfutée, avant 5-6 ans, si la correspondance spatiale terme à terme n'est pas respectée. La conservation de la quantité comptée (avec écran), appelée par Gréco conservation de la *quotité*, serait l'indice d'un niveau fruste de cardinalité si elle n'est pas assortie de la conservation de la quantité en dehors du comptage. Sans cela le cardinal ne serait qu'un label de surface. Conservation de la quotité, de la quantité, compréhension de l'ordre serial et de l'inclusion hiérarchique des nombres devraient être nécessairement inclues dans une évaluation des compétences de l'enfant.

1.2 L'innéisme de Gelman

La théorie est bien connue et je n'en rappelle ici que les traits spécifiquement intéressants pour l'exposé.

De nombreuses observations ont fait état de connaissances «mathématiques» diverses chez le jeune enfant. En témoignent les situations où certaines contraintes liées à l'exécution et au contrôle de l'action sont levées. Un exemple est celui où l'on demande à l'enfant d'évaluer le caractère correct ou incorrect de procédures de comptage effectuées par une poupée : à 3-4 ans la majorité des enfants distinguent parfaitement un comptage correct (avec ou sans pseudo-erreurs) d'un comptage erroné (Gelman et Meck, 1983). A cet âge, les connaissances «mathématiques» des jeunes enfants semblent dépasser d'assez loin celles qui se manifestent lors de situations plus contraignantes. Des observations de ce genre conduisent Gelman à postuler l'existence de compétences implicites dont l'enfant disposerait dès son plus jeune âge. Ces compétences sont décrites sous forme de cinq principes rapportés dans le tableau 1. Il s'agit là de capacités potentielles dont un ensemble assez impressionnant de recherches expérimentales tendrait à confirmer l'existence (voir Gelman, 1982; Gelman et Meck, 1983, 1991...).

Cependant, de nombreuses données empiriques mettent en doute l'existence de principes définis aussi précisément que ceux de Gelman. Un bon exemple en est la recherche de Briars et Siegler (1984) concernant le principe de correspondance mot-nombre/objet et le principe de non pertinence de l'ordre (principes 2 et 5 du tableau 1) qui porte sur des enfants de 3,4 et 5 ans. La situation adaptée de Gelman et Meck (1983) est celle de la poupée qui compte en utilisant trois types de procédures : 1/ le comptage erroné avec violation de la correspondance mot-nombre/objet; 2/ le comptage correct mais inhabituel (comptage d'éléments non adjacents, comptage de droite à gauche ou à partir de la rangée); 3/ le comptage standard de gauche à droite. Les enfants doivent juger du caractère correct ou incorrect des procédures. Les résultats révèlent que la correspondance mot-nombre/objet n'est vraiment acceptée qu'à 4 ans. La compréhension de la nature essentielle de cette règle

se développerait entre 3 et 5 ans sans être encore complètement atteinte à 5 ans. A cet âge, en effet, les enfants rejettent encore, dans leur majorité, les procédures correctes mais inhabituelles (comptage des éléments non adjacents par exemple). Par ailleurs, Briars et Siegler testent l'hypothèse d'une antériorité du comptage correct sur la règle de correspondance. Les résultats observés montrent que 11 des 16 enfants de 3 ans de l'échantillon comptent eux-mêmes correctement sans pour autant rejeter de manière stable les violations de la règle de correspondance effectuées par la poupée tandis que tous les enfants qui les rejettent (n=5) comptent correctement. Le comptage «social» semble bien précéder l'utilisation et la compréhension des principes répertoriés par Gelman (voir aussi Baroody, 1991).

TABLEAU 1

Les principes fondamentaux du comptage selon Gelman et Gallistel (1978)

1. *Principe d'ordre stable* : les mots-nombres doivent être engendrés dans le même ordre à chaque comptage.
2. *Principe de stricte correspondance terme à terme* : chaque élément d'une collection doit être désigné par un mot-nombre et un seul.
3. *Principe cardinal* : le mot-nombre qui désigne le dernier élément d'une collection représente le nombre total d'éléments.
4. *Principe d'abstraction* : seules sont abstraites, des éléments comptés, leurs caractéristiques d'entités distinctes
5. *Principe de non pertinence de l'ordre* : l'ordre dans lequel les éléments d'une collection sont énumérés n'affecte pas le résultat du comptage à condition que le principe de correspondance terme à terme soit respecté

Le principe de cardinalité a fait également l'objet de vives controverses. Réfutant le subitziting, ou aperception globale de la numérosité de petites collections, Gelman et Meck (1991) créditent de l'utilisation du principe cardinal (principe 3 du tableau 2)»...les enfants qui se servent d'un seul terme pour évoquer -soit en réponse à une question, soit spontanément- la numérosité de la collection» (ibid., p.227). Sur de petites collections de'3 et 4 éléments, Gelman et Meck observent, avec ces critères, l'application du principe cardinal chez respectivement 7 et 8 sur 12 enfants de 2 et 3 ans. Mais selon Fuson (1991) et Sophian (1991), le comptage ne débouche pas d'emblée sur un résultat cardinal. Il s'agirait d'une imitation de la conduite adulte. La «règle du dernier mot» serait apprise procéduralement avant la compréhension de ses implications cardinales. Comptage et cardinalité seraient dissociés et ne s'intégreraient que progressivement. Une récente recherche de Grégoire et Van Nieuwenhoven (1993, voir aussi ci-dessous) indique que les principes de Gelman ne sont pas encore réellement intégrés à 5 et 6 ans.

En dépit des critiques nombreuses et d'observations, qui conduisent pour le moins à la moduler, la perspective de Gelman reste intéressante à dou-

ble titre. Tout d'abord elle peut expliquer les différences inter et intra-individuelles. A partir du moment où des compétences bien définies existent, les variations pourraient renvoyer à deux sources d'erreurs mentionnées par Gelman et Meck (1991, p. 220) : «...(a) une *compétence procédurale* inadéquate au niveau de l'établissement des plans d'actions; (b) une *compétence interprétative limitée...* » (compréhension de la consigne, etc.). Le second intérêt réside dans l'essor donné à l'étude des activités numériques précoces et particulièrement à celle du comptage.

Quelles que soient les critiques, dont le bien fondé paraît certain, l'apport de Gelman réside dans l'accent mis sur des principes, bien déterminés et différenciés, que l'enfant qui apprend à compter doit intégrer. D'où l'importance d'évaluer leur niveau d'élaboration. Le principe de cardinalité renvoie-t-il, par exemple, à la même organisation cognitive (représentation et opération) entre 4 et 8-9 ans ?

1.3 Le constructivisme de Fuson et de Steffe

Deux activités doivent être coordonnées lors du dénombrement des éléments d'une collection : la production d'une série ordonnée de mots-nombres et l'application de chacun de ces mots-nombres à un seul des éléments de la collection, en évitant les omissions et les répétitions. Si cette coordination se développe progressivement entre 2 et 6 ans pour la numération élémentaire, il faut attendre 8-9 ans et plus pour qu'elle soit effectivement maîtrisée. A partir d'observations multiples et précises, Fuson (Fuson, 1988, Fuson 1991) distingue quatre niveaux d'élaboration de la chaîne numérique qui correspondent aux étapes mises en évidence par Steffe (1991). Ces niveaux, répertoriés dans le Tableau 2, sont brièvement décrits ci-dessous.

TABLEAU 2

La chaîne numérique verbale : les étapes du développement (d'après Fuson, 1991 et Steffe, 1991)

Niveau 1	
– le «chapelet»	Fuson
– le schème de comptage perceptif	Steffe
Niveau 2	
– la chaîne insécable	Fuson
– le schème de comptage figuratif	Steffe
Niveau 3	
– la chaîne sécable	Fuson
– la suite initiale des nombres	Steffe
Niveau 4	
– la chaîne terminale : suite numérique, emboîtée, sériée, cardinalisée et «unitisée»	Fuson
– la suite tacitement puis explicitement emboîtée	Steffe

1. *Le niveau du chapelet.* L'enfant produit une suite compacte de désignations sans signification numérique : un-deux-trois... qui se produit aussi dans les comptines. Elle conduit à l'appréhension précoce de petites numérosités basées sur des configurations perceptives (*subitizing*, schème perceptif de comptage).
2. *La chaîne insécable*, figurative. La caractéristique de cette étape est l'incapacité de compter à partir d'un nombre donné : l'enfant doit revenir à *un*, la chaîne est *insécable*. Selon Steffe (1991, p.123), *l'usage d'une suite de mots-nombres* renvoie à un enregistrement de l'aspect kinesthésique (patterns digitaux) ou auditif. En dépit de cette restriction, cette phase, qui peut se développer jusque 5-6 ans selon les enfants, est celle du *mot-nombre signifiant le comptage* par opposition à la précédente où *le mot-nombre signifie la configuration*. La capacité notable qui émerge est celle du *comptage jusqu'à un nombre donné* (vers 4 ans).
3. *La chaîne sécable.* A ce niveau les enfants comptent à partir de n'importe quel nombre donné et de n'importe quel nombre à un autre. Le comptage à rebours se développe également ainsi que la capacité à répondre à la question «Qu'est-ce *qui vient avant n ?*» Le comptage est *contrôlé* parce que, selon Steffe (1991, p.126), les éléments de la série renvoient à l'intériorisation des opérations réalisées lors des comptages perceptifs et figuratifs. Le nombre prend alors le statut de symbole dans la *suite initiale* progressivement arithmétisée.
4. *La chaîne terminale.* Une nouvelle habileté émerge : compter n à partir de x dans les deux sens, ce qui implique l'énumération de la suite numérique en même temps que la conservation des nombres déjà émis. La difficulté croît avec le comptage à rebours et la taille de n, l'influence de ce dernier facteur disparaissant avec la scolarité. Au plus haut niveau d'élaboration, la chaîne devient bidirectionnelle, «emboîtée, *sériée, cardinalisée et unitisée*».

On rejoint à la dernière étape la conception piagétienne du nombre, synthèse de la classe et de la série (1<(1+1)<(1+1+1)... etc.). Mais les relations logiques sous-jacentes sont inférées des habiletés d'énumération et de comptage et non des performances réalisées à des épreuves où classification, sériation, et/ou conservation, sont explicitement impliquées. On peut certes penser que le comptage à rebours et le comptage de n à partir de x dans les deux sens sont les manifestations d'une certaine compréhension des propriétés de la séquence numérique. Mais la généralisation de ces conduites est relativement tardive, vers 9 ans selon Fuson et al. (1982, p.70). Des faits expérimentaux nombreux ont indiqué que la maîtrise opératoire de l'inclusion des classes et de la sériation (transitivité) ne survient qu'à partir de 10-11 ans (voir à ce sujet, pour une revue de question, Bideaud 1988; Bideaud et Houdé, 1989; Leiser et Gillièron, 1990). S'agissant alors du nombre, deux éventualités se présentent. Ou l'enfant peut-être crédité dès 7-8-9 ans, de la compréhen-

sion des propriétés logiques de la chaîne numérique (chaîne bidirectionnelle, emboîtée, sériée) et alors il faut : 1) le prouver et 2) si c'est le cas, interpréter cette avance en termes de facilitation entraînée par le caractère d'emblée symbolique, abstrait du nombre. Ou bien l'enfant ne peut être crédité de cette maîtrise avant 10-11 ans (cas de l'inclusion et de la sériation) et on doit alors en tenir compte à la fois dans les évaluations et dans le choix des situations d'apprentissages. Dans ce deuxième cas, une autre interprétation pourrait être donnée : un niveau plus restreint de compréhension de la logique sous-jacente à la chaîne numérique pourrait s'avérer un pré-requis suffisant pour l'apprentissage des algorithmes de l'arithmétique élémentaire. Mais alors où se situerait la dernière étape et la maîtrise de la chaîne numérique ? Et pour quels progrès serait-elle nécessaire ?

1.4 Que nous apportent les théories néo-structuralistes ?

Le néo-structuralisme vise explicitement à articuler les apports du structuralisme genevois et du cognitivisme anglo-saxon dans le but de pallier les insuffisances du cadre explicatif piagétien. Il n'y a pas lieu ici de passer en revue ces théories et d'en faire le bilan (voir à ce sujet Ribeaupierre, 1983 ; Bideaud et Houdé, 1993 ; Bideaud, Houdé et Pédinielli, 1993). Ce sont des explications très générales qui tentent de spécifier les étapes du développement tant au niveau du fonctionnement *in situ* qu'à celui des structurations cognitives sous-jacentes. Aucune d'elles ne prend pour cible la genèse du nombre et les acquisitions scolaires à l'exception toutefois de Case qui aborde, à la périphérie de ce domaine, la lecture de l'heure et la valeur et l'usage de la monnaie. On ne retiendra ici que certains concepts-clés, heuristiques, des perspectives de Pascual-Leone, de Case et de Fischer, susceptibles d'aider à la compréhension du fonctionnement et du développement de l'enfant dans les situations «mathématiques».

1.4.1 *Pascual-Leone : apprentissages et système modulaire de l'attention mentale*

Pascual-Leone reprend la distinction piagétienne entre apprentissage des contenus, c'est-à-dire des qualités des objets, (opérateur C) et apprentissage structural, c'est-à-dire des relations et des systèmes de relations, (opérateur L, logique). Mais à la différence de Piaget, pour toutes les acquisitions qui ne relèvent pas du seul conditionnement (opérateur C), l'opérateur «logique», L, n'entre jamais seul en action, mais sous deux formes : LC et LM. L'opérateur M renvoie à l'énergie mentale spécifiée par le maximum de schèmes ou connaissances représentées que le sujet peut activer dans une opération mentale, maximum qui croît avec l'âge. L'apprentissage LC correspond à la fusion graduelle de schèmes - connaissances représentées de tous genres - par confrontation répétée du sujet avec les mêmes régularités. Il s'agit d'un apprentissage lent, cumulatif et tacite, qui génère des structures généralement automatisées et fortement intégrées dans leurs contextes d'origine. In-

versement, l'apprentissage LM, suscite, grâce au poids de M, une articulation rapide, *associée à un effort cognitif*, des schèmes (connaissances et opérations) pertinents. Cette rapidité entraîne une autonomie des structures générées qui sont alors applicables dans divers contextes. *L'apprentissage LC génère le répertoire des connaissances «expérientielles» tandis que LM génère celui des connaissances rationnelles.*

Dans le cadre de sa théorie et en liaison avec d'autres opérateurs (I=inhibition ; F=facteur de champ qui filtre ou maximise les informations), Pascual-Leone distingue deux modes de fonctionnement possible, l'un automatisé, l'autre, attentionnel. Le mode attentionnel a été récemment précisé dans un système à quatre modules correspondant aux opérateurs M, I et F et aux schèmes exécutifs du sujet qui sont des programmes de calcul. En référence à ces modules et à leur interaction, Pascual-Leone décrit différentes stratégies attentionnelles correspondant à des problèmes spécifiques. A noter l'importance donnée à l'opérateur d'inhibition, I (inhibition des schèmes «dangereux» (par exemple, dans les épreuves de conservation) ou non pertinents, ou moins pertinents).

On peut appliquer cette conceptualisation à l'apprentissage de la chaîne numérique et à celui de la genèse des algorithmes (addition, soustraction). Selon J.P. Fischer (1993a, p.54) le comptage, entre 2-3 et 5-6 ans, est une connaissance procédurale, résultat d'un apprentissage lent et graduel, «expérientiel», qui «nécessite une pratique considérable». De ce fait, il peut se révéler très contextualisé et donc plus difficilement «exportable» dans des situations variées. Mais alors quand et comment s'effectue l'apprentissage *rationnel* qui doit conduire à l'assimilation des propriétés logiques de la série numérique ? On touche du doigt ici la nécessité d'une estimation de la qualité des acquisitions en maternelle et en première année primaire pour une meilleure compréhension des difficultés et des erreurs. La contextualisation trop poussée assortie d'un défaut de capacité d'inhibition est aussi, sans aucun doute, une cause d'erreurs.

1.4.2 *Case et les structures de contrôle*

Dans la perspective de Case, toute situation d'interaction entre le sujet et l'environnement constitue un problème à résoudre. Le développement cognitif devient celui de *Structures de Contrôle Exécutif* (SCE) ou séquences organisées de schèmes à trois composantes : la *représentation du problème* dans lequel le sujet est impliqué, l'*objectif* doté le plus souvent d'une charge «affective» et les *stratégies* qui permettent de passer de la représentation à l'objectif. La genèse des SCE est jalonnée par une séquence de 4 stades (sensori-moteur, relationnel, dimensionnel et vectoriel) qui correspondent grosso modo aux stades piagétiens et qui se différencient par le type d'éléments mentaux que traitent les SCE (objets perçus et activités motrices, relations entre activités, dimensions du 1er ordre et dimensions du 2ème ordre (rapports de dimensions). Chaque stade comporte trois sous-stades distingués par le nombre de cibles, de 2 à 4, impliqués dans la structure

de contrôle (objectifs). Cette augmentation est liée à celle de la capacité de mémoire à court terme, elle-même fonction de l'augmentation de l'efficience opératoire. L'efficience opératoire, par ailleurs, s'accroît au cours d'un stade sous l'action conjuguée de l'exercice et de la maturation.

A ce modèle des structures de contrôle exécutif, Case a récemment ajouté la notion de *Structure Conceptuelle Centrale* (SCC). Il s'agit de réseaux de concepts interreliés, de nature logico-mathématique ou sociale, qui correspondent aux répertoires de connaissances transituationnelles. Ce sont eux qui *donnent sens* aux situations problèmes rencontrées. Une de ces SCC est relative au nombre. Ainsi à tout moment de son développement l'enfant dispose t-il de connaissances générales, plus ou moins élaborées et plus ou moins contextualisées, qui «dirigent» ses structures de résolution. L'élaboration de ces connaissances dépend évidemment du type d'apprentissage, au sens très général du terme, qui est à leur source. On retrouve ici la question posée par les opérateurs de Pascual-Leone.

Le recours à la mémoire à court terme est intéressant. S'agissant de la séquence numérique, c'est sans aucun doute l'augmentation de cette capacité qui permet à l'enfant de 4 ans (dernier sous-stade du stade relationnel) de compter jusqu'à un nombre donné (conservation de ce nombre pendant le comptage). C'est elle aussi qui permet à l'enfant de 6-8 ans (1er et 2ème sous-stades du stade dimensionnel, qui implique la compéhension de relations entre éléments) de compter n à partir de x dans les deux sens.

1.4.3 *Kurt Fischer et l'enfant-en-contexte*

Le concept central de la théorie de K. Fischer est celui de structures de *skills*, ou structures d'habiletés de résolution de problème, qui se développent, de la naissance à 30 ans, en quatre étages : réflexe, sensori-moteur, représentationnel et abstrait. A chaque étage, le sujet se développe selon quatre niveaux de complexité, qui correspondent, comme chez Case, aux types d'éléments mentaux traités par les structures de *skills* (unités sensori-motrices, appariement, systèmes, systèmes de systèmes).

La spécificité de K. Fischer réside dans le fait que les étages, les niveaux et les processus de transition sont toujours ceux d'un *sujet-en-contexte* et que les décalages des performances individuelles, suivant les supports contextuels, *sont la règle du développement cognitif.* D'où l'introduction, par K. Fischer, de la distinction entre *niveau optimal et niveau fonctionnel suboptimal.* Le niveau optimal correspond à la limite supérieure de complexité du répertoire des habiletés dans un domaine donné. Il est atteint dans une situation problème donnée, lorsque le support contextuel est un *support à haut niveau de performance* : support contextuel optimal. En revanche, lorsque le support est suboptimal le sujet fonctionne à un niveau inférieur, dit «fonctionnel».

Il semble que cette distinction entre niveaux de fonctionnement puisse être des plus intéressantes pour l'évaluation des capacités

«mathématiques» de l'enfant. En jouant sur le contexte - et c'est ce que fait Gelman avec son paradigme de la poupée qui compte - on peut évaluer à la fois le niveau «fonctionnel» et, en graduant les difficultés, le niveau optimal. Il faut souligner qu'avec Fischer, plus encore que chez les autres néo-structuralistes, le contexte des acquisitions joue plus qu'un rôle primordial : *il n'y a pas de sujet sans contexte et les deux sont indissociables.*

Cet exposé, drastiquement raccourci de quelques concepts-clés des théoriciens néo-structuralistes, n'apporte pas des éléments directs à la recherche d'une évaluation diagnostique fine des capacités numériques de l'enfant. Mais il montre une convergence sur certains points cruciaux dont l'évaluation doit tenir compte : type des apprentissages explicites et implicites, scolaires et extra-scolaires, rôle des contextes d'acquisition et d'évaluation, importance de l'attention et, partant, des capacités d'inhibition.

2. L'ÉVALUATION DES CAPACITÉS «MATHÉMATIQUES» DES JEUNES ENFANTS : LES DONNÉES EMPIRIQUES

Peu de chercheurs ont exploré systématiquement les capacités numériques des jeunes enfants. Trois études ont été ici retenues pour une analyse comparative. Deux d'entre elles, celle de Ginsburg et Russell (1981), et celle de Grégoire et Vanniewenhoven (1993), sont remarquables à plusieurs titres : la première pour l'empan des conduites observées, la seconde pour la précision méthodologique, la finesse de l'analyse et l'importance de la population concernée. Le but de la troisième n'est pas, à proprement parler, un bilan des capacités. Mais dans la mesure où elle introduit des apprentissages, elle montre ce que pourrait être une *évaluation dynamique.*

2.1 La recherche de Ginsburg et Russell (1981)

De nombreuses observations font ressortir un échec scolaire important, en mathématique, chez les enfants de milieux défavorisés. L'explication généralement avancée attribue cet échec à un déficit des stimulations de l'environnement : celui-ci n'offrirait pas aux enfants les expériences nécessaires au développement des pré-requis sur lesquels se fonde l'apprentissage mathématique à l'école. Ginsburg et Russell explorent l'interaction possible entre le type de population (blanche et noire), la classe sociale (niveau bas et moyen), le statut familial (monoparental ou normal) et les connaissances mathématiques pré-scolaires. L'objectif est de vérifier - outre l'étude de ces interactions - si l'échec scolaire peut être prédit à partir de différences dans les connaissances mathématiques précoces, pré-scolaires.

L'étude principale porte sur 144 enfants dont la moitié fréquente des garderies (a.m.= 4;8) et l'autre moitié les jardins d'enfants (a.m.= 5;8). La moitié des enfants de chacune des populations appartient à un milieu très défavo-

risé et l'autre à un milieu moyen, l'indicateur choisi pour opérationnaliser la variable étant le *Hollinsgshead Scale* (index des positions sociales). A chacun de ces niveaux sociaux, la moitié des enfants appartient à la race noire et l'autre moitié à la race blanche. La variable «statut familial» est équilibrée dans chacun des sous-groupes. Au cours de trois séances, les enfants sont soumis individuellement à 17 tâches qui sont répertoriées dans la colonne de gauche du tableau 3. Ces tâches recouvrent le champ possible des connaissances «logico-mathématiques» précoces. On peut les regrouper sous trois rubriques.

I Évaluation des aspects non scolaires et non numériques de la pensée «mathématique».

1. Perception du «plus» : l'enfant doit indiquer parmi deux ensembles, à configuration variée, celui qui en a le plus.
2. Sériation de 7 baguettes.
3. Conservation numérique classique.
4. Addition-ajout d'éléments et compréhension du résultat.
5. Jugement d'équivalence de deux collections disposées différemment.

II Évaluation au moyen de tâches non scolaires qui peuvent être résolus par des procédures de comptage.

6. Comptage sans support (borne fixée à 50).
7. Dénombrement de plusieurs collections différentes (6 collections de 3 à 19 objets).
8. Cardinalité (avec les collections précédentes).
9. Indifférence de l'ordre (4 items).
10. Principe d'abstraction : comptage d'une collection hétérogène.
11. Détermination du plus grand de deux nombres donnés oralement.
12. Règle de l'unité : l'ajout ou le retrait d'une unité à un ensemble d'objets accroît ou décroît d'une unité la séquence de comptage de cet ensemble.
13. Application de la règle 12.
14. Addition-calcul : résolution avec ou sans supports d'additions de chiffres qui vont de 2 à 7.

III Évaluation de la représentation du nombre.

15. Évaluation d'une représentation quelconque du nombre (traits, figures, écriture des chiffres).
16. Lecture de nombres.
17. Écriture de nombres.

Ginsburg et Russell observent un effet massif de l'âge sur toutes les épreuves sauf pour l'épreuve «principe d'abstraction» réussie par les 144 enfants (cf. tableau 3). A 4 ans, la variable statut social introduit une différence significative dans les épreuves de conservation, d'équivalence numérique de quantités égales diversement disposées et dans la comparaison de grandeur de deux nombres énoncés. Cet différence disparaît à 5 ans (effet de l'âge con-

jugué avec le programme du jardin d'enfant) pour les deux dernières épreuves. La variable population (noire *vs* blanche) exerce un effet sur la sériation et la règle de l'unité.

TABLEAU 3

Scores moyens observés en fonction de l'âge aux différentes épreuves de Ginsburg et Russell (1981)

			Scores maxi.	4 ans (n=72)	5 ans (n=72)
1	Perception du « plus »	*a*	**6**	5,17	5,69
		b	**6**	4,80	5,15
		c	**6**	2,21	2,50
2	Sériation		**6**	2,55	4,32
3	Conservation		**6**	2.44	3.57
4	Additions (ajouts)		**4**	2.68	3.40
5	Equivalence		**12**	3.15	4.28
6	Mots nombres		**50**	24.66	36.94
7	Dénombrement		**6**	1.60	2.27
8	Cardinalité		**6**	4.50	5.70
9	Indifférence de l'ordre		**4**	2.32	2.95
10	Principe d'abstraction		**1**	1	1
11	Nombre plus grand		**10**	7.24	8.89
12	Règle de l'unité		**12**	9.19	11.40
13	Application de la règle 12		**8**	5.63	7.58
14	Addition (calcul)		**6**	2.54	3.91
15	Représentation du nombre		**3**	2.83	2.36
16	Lecture des nombres		**4**	3.12	3.72
17	Ecriture des nombres		**8**	2.33	5.81

Ces résultats sont intéressants à double titre. Tout d'abord ils permettent de dresser un inventaire à la fois large et précis des connaissances logico-mathématiques des jeunes enfants. Elles se révèlent consistantes et variées avec un effet puissant de l'âge. A 5 ans, la majorité des enfants réussissent la sériation des baguettes, connaissent et appliquent la règle de l'unité, connaissent la règle du dernier mot-nombre (objets en ordre). Les moyennes élevées en témoignent. Et si l'effet de l'âge est puissant, l'observation du tableau 3 témoignent de connaissances certaines. Selon Ginsburg et Russell, la connais-

sance mathématique «extra-scolaire» se développerait de manière précoce et robuste. En second lieu, ces résultats démontrent que, mises à part la conservation et l'équivalence, les enfants arrivent à l'école primaire, quel que soit leur statut social, avec des connaissances déjà nombreuses et importantes. Mais alors pourquoi les enfants des milieux défavorisés, et dans ces milieux plus particulièrement les enfants noirs, échouent-ils aussi massivement lorsqu'il s'agit des mathématiques scolaires ? Les expectatives ne sont pas en cause. Selon Ginsburg et Russell, aux âges observés et même à la fin de la première année scolaire, les enfants pauvres noirs et blancs s'attendent à réussir en mathématique au long de leur cursus. Les auteurs invoquent alors la disparition de cette motivation en raison de difficultés liées au statut social et l'inadéquation de l'enseignement scolaire.

C'est ce facteur que nous retenons ici et pour deux raisons. La première raison tient au caractère intrinsèque de l'enseignement mathématique qui ne prend pas très souvent en compte les significations que le jeune enfant attache au nombre avant et hors de l'école (voir Sophian, 1991). La seconde tient à la fois à certaines caractéristiques des milieux sociaux et à l'enseignement scolaire. Dans les milieux défavorisés l'enfant est confronté très tôt à des problèmes pratiques de calcul. Il intègre de ce fait davantage de schémas et des routines rigides qui peuvent freiner des apprentissages scolaires plus abstraits. Les résultats de Ginsburg et Russell nous donnent les performances moyennes : les procédures sous-jacentes ne sont pas explicitées. Il se dégage donc encore ici à la fois la nécessité d'une évaluation plus fidèle susceptible d'orienter des apprentissages adaptés et le rôle des contextes plus ou moins favorables d'acquisition du point de vue de leur rigidité ou de leur généralisation possible.

2.2 La recherche de Grégoire et Van Nieuwenhoven (1993)

Les auteurs se proposent d'évaluer les capacités de comptage en troisième maternelle et en première primaire. Pour ce faire, ils construisent des épreuves susceptibles d'appréhender le plus précisément possible un fonctionnement correspondant aux principes de Gelman. La population de l'expérience comprend 439 enfants de 5 ans (troisième année maternelle) et 103 enfants de 6 ans (première primaire).

Le matériel est constitué de deux ensembles de bouchons : l'un de 12 bouchons de plastique transparent ; l'autre de 6 bouchons de plastique bleu. Pour les trois premiers principes de Gelman (ordre stable, correspondance mot-nombre et dernier mot-nombre : cardinalité de Gelman), deux épreuves sont construites : production et détection. Pour les deux autres principes (abstraction et indifférence de l'ordre), une seule épreuve : la production. L'ensemble de ces épreuves est décrit ci-dessous.

1. Ordre stable

1.1. Production
- compter le plus loin.
- compter jusqu'à 9.
- compter à partir de 3.
- compter de 5 à 9.
- compter à rebours à partir de 7.

1.2. Détection
L'expérimentateur compte en commettant des erreurs et des pseudo-erreurs (4) types d'erreur).

2. Principe de correspondance

2.1. Production : 2 items
- l'enfant doit compter les 12 bouchons d'abord alignés, puis dispersés.
- même tâche avec changement de la disposition.

2.2 Détection : 2 items
- comptage opéré par l'expérimentateur en violant le principe de correspondance terme à terme.
- variation sur la situation précédente.

3. Dernier mot-nombre (cardinalité)

3.1. Production
Après avoir compté les bouchons alignés, l'enfant doit répondre à la question : combien y en a t-il en tout ?

3.2 Détection
L'expérimentateur compte une fois correctement puis en faisant une erreur de double comptage. L'enfant doit répondre à la même question que plus haut.

4. Principe d'abstraction
L'enfant doit compter les éléments hétérogènes (forme, taille, couleur) d'un ensemble d'objets.

5. Non pertinence de l'ordre
Tâche A. L'enfant compte 12 bouchons alignés. Après son comptage, on lui demande s'il obtiendrait le même nombre (donné par l'enfant) s'il commençait son comptage par l'autre extrémité.

Tâche B. Même procédure avec les bouchons dispersés.

Les pourcentages de réussite (effectif) sont donnés dans le tableau 4 pour toutes ces épreuves.

TABLEAU 4

Performances (%) observées aux épreuves de Grégoire et Van Nieuwenhoven (1993)

Epreuves	5 ans	6 ans
1. Ordre stable		
1.1. Production		
compter le plus loin	25	43[a]
jusqu'à 9	83	98
à partir de 3	48	57
de 5 à 9	46	70
à rebours depuis 7	27	75
1.2. Détection erreurs		
comptage aléatoire	92	92
inversion	88	95
omission	87	88
répétition	92	94
2. Principe de correspondance		
2.1. P. bouchons alignés	79	93
dispersés	66	83
2.2. D. bouchons compté 2 fois	77	85
sauté	81	84
3. Dernier mot-nombre		
3.1. P. bouchons alignés	64	90
dispersés	62	89
3.2. D. compt. exp. correct	73	86
incorrect	70	82
4. Principe d'abstraction	réussite	réussite
5. Principe d'indifférence de l'ordre		
bouchons alignés	67	85
dispersés	65	87

a. La première ligne indique le résultat moyen au comptage. Pour toutes les autres lignes, il s'agit de l'effectif en %.

La comparaison avec les résultats de Ginsburg et Russell concernant les épreuves communes est intéressante bien que mal aisée, l'une donnant des moyennes de performance et l'autre des pourcentages de réussite. Néanmoins quelques points ressortent (on rappelle que l'expérience de Ginsburg et Russell porte sur des enfants de 4 et 5 ans, celle de Grégoire et Van Nieuwenhoven sur des enfants de 5 et 6 ans) :

1. Le comptage d'éléments hétérogènes (principe d'abstraction) est massivement réussi à 4, 5 et 6 ans.
2. L'épreuve du dernier mot-nombre (cardinalité) est bien réussie à 5 ans dans les deux recherches quand il s'agit de la production (pas de détection dans Ginsburg et Russell).

3. L'épreuve «indifférence de l'ordre» paraît atteindre dans les deux cas, à 5 ans, le même niveau de réussite.
4. L'effet de l'âge se fait sentir aussi bien entre 4 et 5 ans qu'entre 5 et 6 ans : il existe bien un développement du respect des principes entre 4 et 6 ans.

La différence des objectifs des deux recherches oriente différemment les conclusions des auteurs. Ginsburg et Russell soulignent l'importance et la variété de compétences précoces. Grégoire et Van Nieuwenhoven sont plus sensibles aux erreurs des enfants de 5 et 6 ans et il faut dire que leurs épreuves de comptage sont bien plus précises et complètes que celles de Ginsburg et Russell. Leur contexte sollicite vraisemblablement le *niveau optimal* défini par K. Fischer tandis que Ginsburg et Russell, avec leur situation comparable, solliciteraient un *niveau fonctionnel.*

Il semble que l'on puisse conclure à l'existence d'un certain niveau de compétence assez bien établi dès 4 ans avec, dans le cas des deux recherches, une notable variation inter-individuelle. Dans les deux cas on doit toutefois souligner qu'il s'agit, dans la plupart des épreuves (comptage) d'une évaluation de *surface* (cf. Jonnaert, 1993), au sens où, à performance égale, rien ne permet de différencier un enfant de 4 ans ou de 5 ans d'un enfant de 6 ans du point de vue des compétences réelles. Or, outre les différences inter-individuelles de performance, il existe des différences à ce niveau. Il semble qu'une étude longitudinale telle que Grégoire et Van Nieuwenhoven la préconisent avec raison pourrait intégrer, outre leurs épreuves si précises, une évaluation de la «logique» sous-jacente : épreuves d'inclusion et de sériation, appliquées à des objets mais aussi à la séquence numérique.

2.3 L'expérience de Clements (1984)

Il s'agit là d'une expérience d'apprentissage. Je l'expose ici pour la raison qu'il m'a toujours semblé que l'entraînement à une certaine notion par telle ou telle méthode peut renseigner sur les processus actuels d'acquisition qui ne sauraient être étrangers aux organisations actuelles de l'enfant. Les effets d'un entraînement spécifique peuvent être considérés, pour le moins, comme des indicateurs du développement actuel de la notion, plus sensibles que les réponses aux seules épreuves d'évaluation.

Clements (1984) compare, sur des enfants âgés de 4 ans (m.a.=4;6 ans), les effets respectifs d'un entraînement aux activités numériques (N) et d'un entraînement aux opérations logiques (L) (3 séances par semaines de 20 mn. chacune pendant 8 semaines : soit 24 séances pour l'un et l'autre). L'entraînement N porte sur des activités numériques diverses : stratégies de comptage, recherche d'un ensemble de numérosité donnée parmi sept ensembles de numérosités variables, production d'un ensemble de numérosité donnée parmi 20 objets, etc. L'entraînement L est basé sur des exercices de classification et de sériation simples et multiples. Le pré et le post-test com-

portent 10 épreuves numériques, dont la conservation classique (score total possible : 59 points) et 11 épreuves logiques (score total : 50 points). La comparaison des gains observés entre pré et post-test dans les deux groupes expérimentaux et dans un groupe contrôle, C, indique : 1) Dans les deux domaines, numérique et logique, les deux groupes N et L réalisent des gains significatifs ; 2) Les enfants N réalisent des gains plus importants lorsqu'il s'agit des concepts numériques : 27,67 points gagnés contre 11,06 pour les enfants L (différence moyenne entre pré et post-test). Ces mêmes enfants font également des progrès dans le domaine logique du même ordre que ceux des enfants L (respectivement 19,93 et 19,94) ; 3) Le transfert «nombre-logique» est de loin plus important que le transfert «logique-nombre», bien que celui-ci soit certain ; 4) Concernant l'épreuve classique de conservation de l'équivalence (items de 7 et 9 éléments, score maximal = 4), absente des deux entraînements, le groupe N passe, entre pré et post-test, de 1,40 à 3,80 (score moyen), les enfants de L, de 0,87 à 1,73 et ceux de C restent à 1,30. Concernant le comptage de différents ensembles en ordre et en désordre de 6 à 12 éléments, les moyennes passent de 5,27 à 8,73 en N, de 5,13 à 7,07 en L (score maximal 9). Dans l'épreuve où les enfants devaient donner le nombre correct d'éléments d'un ensemble de 7 ou 8 objets, dont on croissait ou décroissait de 1 ou 2 le nombre, les moyennes passent de 1,33 à 5,87 dans le groupe N, de 1,40 à 3,40 en L et de 1,73 à 2,40 en C. On rappelle qu'il s'agit d'enfants de 4 ans.

Au vu des résultats, Clements écrit : «Le transfert démontré pour chacun des entraînements revêt une signification psychologique et pédagogique. Si, à l'évidence, les classes, les sériations et le nombre sont interdépendants, l'apprentissage numérique reste cependant prioritaire» (ibid., p.774). Ces résultats et interprétations ne discréditent pas la théorie piagétienne mais accordent aux activités numériques un rôle déterminant (position de Fuson, 1991). Ils soulignent aussi l'intérêt d'une *évaluation dynamique* qui réduirait l'écart entre le niveau actuel et le niveau potentiel de l'enfant, ou, selon K. Fischer, entre un niveau fonctionnel et un niveau optimal.

3. CONTEXTES D'UTILISATION, D'ÉVALUATION ET D'APPRENTISSAGE

Sans aucun doute l'enfant qui entre à l'école primaire possède déjà des connaissances numériques considérables. L'exposé précédent en témoigne (pour un relevé plus exhaustif de ces savoirs et savoir-faire, on peut se reporter à Young-Loveridge, 1987). Quelle que soit la perspective sous-jacente et quelles que soient les interprétations suscitées, la construction du nombre se révèle progressive et complexe. De nombreux processus y concourent qui touchent à la quantification, aux capacités de mise en relation (correspondance, ordre, emboîtement), à la mémoire à court et à long terme, et même aux métaconnaissances. Concernant ce dernier point on sait, par exemple, qu'une

réflexion suscitée sur l'exactitude de leurs réponses conduit des enfants de 4 ans à rechercher un indice fiable de jugement de conservation - en l'occurrence, la numérosité (Michie, 1984). De nombreux facteurs interviennent aussi. Les facteurs linguistiques jouent au niveau de la chaîne numérique verbale dont ils conditionnent en partie l'intégration progressive. La figure 1 tend à donner une représentation à l'évidence approximative et partielle des domaines concernés par la construction du nombre. Y sont ajoutés, en toile de fond, les contextes pragmatiques (utilisation du nombre dans la vie quotidienne) et normatifs (utilisation plus abstraite du nombre dans l'apprentissage scolaire) d'utilisation et donc d'acquisition et d'évaluation.

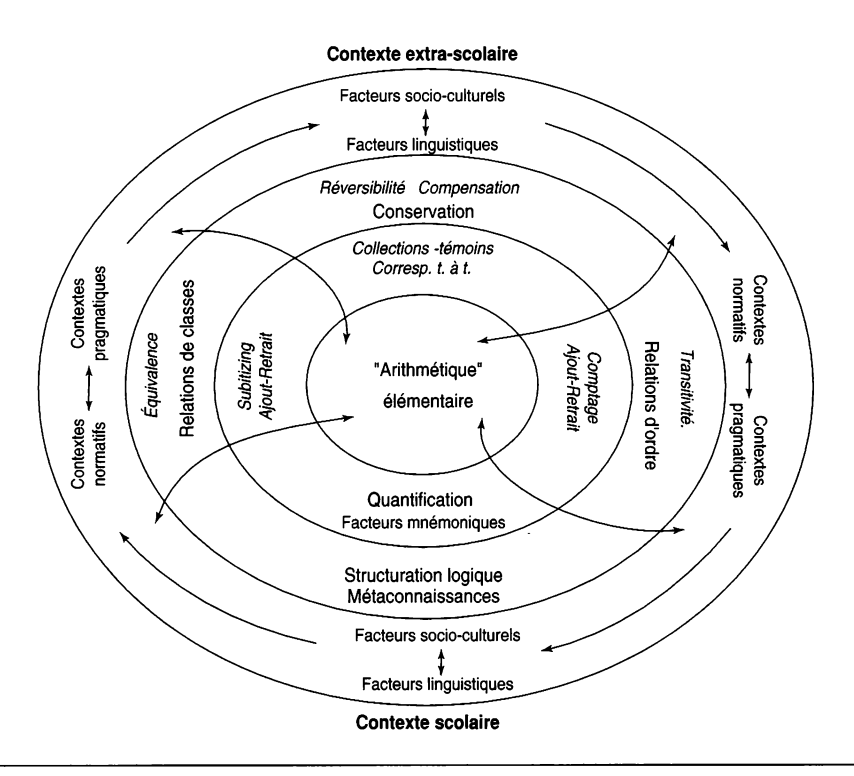


FIGURE 1

Les domaines concernés par la construction du nombre (d'après Bideaud et Vilette, 1995).

Les nombres sont des outils culturels à usages divers intégrés dans des contextes très différents. S'agissant de quantités discontinues, ils indiquent soit la totalité d'un ensemble («Pierre a trois frères et deux soeurs»), soit le rang d'un élément dans une suite («Pierre est le troisième enfant de la famille»). Si la dénotation verbale permet en général de distinguer précisément la cardinalité de l'ordinalité («quatre» et «quatrième» par exemple), de nombreux cas d'utilisation du nombre sont ambigus à cet égard. Dans

l'assertion : «J'habite au 4 de la rue Vauban», le chiffre 4 ne renvoie pas à la cardinalité ; il indique un lieu (la maison de la rue Vauban qui a le numéro 4) et un ordre (la maison qui, dans la rue, est située avant celle qui porte le n°6 et après celle qui porte le n°2). S'agissant de la quantité continue, le nombre devient une «mesure» en indiquant les unités qui correspondent à la quantité en cause : «Je voudrais 3 mètres de ce tissu» ou «Pesez 250 grammes de farine». Hors de toute fonction cardinale, ordinale ou de mesure, le nombre peut aussi être utilisé comme un index discriminatif : c'est le cas des numéros de carte de sécurité sociale, de téléphone, etc. C'est aussi l'un des sens du «4» de la rue Vauban cité plus haut.

Ces quelques exemples montrent à la fois la diversité et la généralité de l'emploi du nombre. Nous vivons dans un monde chiffré et les enfants, interrogés dans nos recherches lors d'exercices de laboratoire, baignent dans ces chiffres, leur donnent des significations, se les approprient pour les utiliser à leur manière. Dans une recherche en cours (Bideaud et Daïf, à paraître), l'évolution des significations des nombres impliqués dans la vie quotidienne est étudiée chez des enfants de 3 à 8 ans. Des enfants de 5 ans à qui l'on demande la signification du numéro accolé au nom de leur rue donnent des réponses très diverses : «c'est le numéro qui est sur la porte» ; «c'est le numéro pour dire au facteur de mettre les lettres» ; «il y a un 7 pour dire que la maison est après le 5». Certains enfants essaient déjà d'expliquer, plus ou moins confusément, l'opposition des numéros, pairs et impairs, de chaque côté de la rue. Les différences sont vraisemblablement extrêmement dépendantes du contexte familial et de la pédagogie parentale d'explication et d'utilisation du nombre. Il n'existe, à notre connaissance, aucune étude systématique des pratiques éducatives familiales dans ce domaine et des différences qu'elles provoquent. Rieunaud (1989, p.57) écrit qu'au cours de ses recherches sur la comptine, il a été surpris d'entendre de nombreux enfants lui dire qu'ils avaient appris à compter «dans la voiture de leurs parents, en comptant les voitures qui passaient». Or, selon Rieunaud, il s'agirait là d'une situation d'apprentissage particulièrement favorable. «L'enfant, en voiture, est un enfant éducable, car il est en situation de curiosité disponible ; éducable d'autant plus que c'est dans un tourbillon d'images qui le fouette et le ravit, qu'on propose un aliment à son attention : repérer des automobiles!» (ibid., p.57).

L'utilisation du comptage, quel que soit son degré d'élaboration, se généralise très vite à des situations de la vie quotidienne bien différentes de celles que l'enfant rencontrera à l'école ou qu'il rencontre s'il fréquente déjà l'école primaire. Dans la vie quotidienne, il s'agit d'un contexte pragmatique qui suscite, dans l'interaction sociale, des procédures finalisées. Les enfants utilisent le nombre très tôt dans des séquences d'interactions sociales pratiques - notamment dans leur rapport avec l'argent : argent de la tirelire compté en vue d'un achat, rémunération de petits services, argent confié pour les courses, etc. L'enfant peut disposer de routines de calcul difficilement exploitables dans un autre contexte. C'est la raison possible des difficultés et des différences observées lors de situations plus abstraites. Une observation de

Carraher et al. (1985) est très instructive à cet égard. Les auteurs citent le cas d'enfants de Récife, vendeurs ambulants de noix de coco. Soucieux de ne pas se tromper dans leur calcul, ce qui entraînerait la perte de leur clientèle, ces enfants se méfient des «trucs» enseignés à l'école. C'est ainsi que la règle apprise selon laquelle pour multiplier un nombre par 10, il faut lui ajouter un zéro, ne leur paraît pas du tout sûre, faute d'une compréhension exacte de ce qu'elle recouvre. Ils opèrent autrement. C'est la cas, par exemple, d'un enfant qui calcule le pris de 10 noix de coco, à 35 centavos pièce, de la manière suivante : (105+105+105)+35. Il a vendu souvent 3 noix, et il sait à coup sûr que 35x3=105. Il sait aussi sûrement que 10=3+3+3+1. D'où le calcul effectué, gage de sécurité. Cet exemple tendrait à montrer qu'une interrogation portant sur les compétences sociales pratiques, quotidiennes, du jeune enfant peut fournir des indices précieux quant à ses capacités réelles et quant aux processus de résolution qu'il engage.

4. CONCLUSION

La construction du nombre est complexe. Son évaluation diagnostique l'est tout autant. Dans cet exposé, j'ai voulu soulever des questions trop souvent éludées et suggérer des pistes pour une évaluation diagnostique la plus exhaustive possible des capacités «mathématiques» des jeunes enfants, compte tenu de certains apports de la psychologie cognitive.

Quatre points, me semble t-il, devraient plus particulièrement retenir l'attention : l'estimation des compétences sous-jacentes aux procédures observées, l'effet des contextes sur la nature des acquisitions, l'importance des capacités générales de résolution de problème et, enfin, la nécessité d'une évaluation dynamique et non seulement statique.

4.1 Structures et procédures

On donne ici au terme «structure» le sens très général d'organisation cognitive sous-jacente, sans statuer sur la nature des liaisons impliquées. Un des apports de la psychologie cognitive constructiviste est d'avoir montré que des procédures de réussite ou d'échec identiques peuvent renvoyer à des organisations très différentes de connaissances spécifiques et/ou plus générales (voir, par exemple, Bideaud et Houdé (1989) pour la catégorisation). C'est l'évidence pour la cardinalité dont le statut développemental est bien différent suivant qu'elle est associée ou non à la conservation des éléments discrets en cause. Peut-on concevoir qu'une cardinalité précoce qui s'accommode d'une variation de la quantité en fonction de transformations perceptives apparentes puisse avoir la même signification que celle qui maintient l'invariance dans les mêmes conditions ?

Il semble bien que les apports théoriques et méthodologiques des deux approches, structurale et fonctionnelle, doivent être pris en compte complémentairement. Dans la pratique, cette utilisation «synthétique» s'avè-

re plutôt difficile. Elle réclame un véritable effort de recherche méthodologique. Mais le jeu en vaut la chandelle et cela ressort avec force du chapitre que Grégoire consacre, dans ce volume, à l'évaluation diagnostique des troubles d'apprentissage en mathématiques.

4.2 Contextes et apprentissages mathématiques

On tend à oublier que l'enfant, tout comme l'adulte d'ailleurs, est toujours un sujet-en-contexte. L'enseignement mathématique à l'école primaire ignore le plus souvent les connaissances antérieures de l'enfant, acquises tout autant à l'école maternelle qu'au sein de l'environnement familier. Le chercheur les ignore aussi qui ne manipule que les aspects extérieurement observables de la performance des sujets. C'est le cas de Siegler et du modèle - très intéressant par ailleurs - qu'il propose pour expliquer les différences intra et inter-individuelles dans la résolution de problèmes d'arithmétique élémentaire (voir Siegler, 1988 pour l'une des principales publication à ce sujet ; voir aussi Reuchlin, 1991, et Lemaire, 1993, pour une analyse précise du modèle). La répartition *a posteriori* effectuée par Siegler entre bons et mauvais élèves, au vu des performances, n'est pas satisfaisante. A ce sujet, Reuchlin écrit (1991, p.209) : « Une utilisation différentielle plus fondamentale consisterait à émettre des prévisions individuelles sur la stratégie employée... et sur les différents aspects des performances réalisées, en assignant, pour chaque individu, une valeur à certaines variables figurant dans le modèle ». Reuchlin évoque alors, parmi les variables possibles, les apprentissages spécifiques préalables. On peut aussi invoquer, outre ces apprentissages, un champ plus large qui englobe les connaissances numériques générales tout autant que les apprentissages spécifiques, en distinguant *les apprentissages scolaires normatifs des apprentissages pragmatiques extra-scolaires*. Dans cette visée, l'inventaire des acquis procéduraux extra-scolaires pourrait s'avérer précieux. Et cela vaut pour les apprentissages scolaires. S'agissant de la multiplication, par exemple, Brissiaud (à paraître), souligne les différences qui peuvent survenir dans sa résolution (mobilité et efficacité) selon qu'a été utilisée, pour l'apprentissage, la table des additions réitérées (1 fois 3, 2 fois 3,...) ou la table classique (3 fois 1, 3 fois 2,...). Il montre aussi que c'est seulement dans une culture de l'écrit que la multiplication peut devenir autre chose qu'une simple addition répetée. On peut alors penser que le poids différent donné à l'oral et à l'écrit puisse introduire, selon les contextes scolaires d'apprentissage, des différences inter-individuelles dans l'organisation des connaissances déclaratives et dans les procédures de traitement, qu'il s'agisse de la multiplication ou de l'addition.

4.3 Les capacités générales de résolution de problème

En dehors des connaissances spécifiques nécessaires, toute situation-problème exige, pour sa résolution, la mise en oeuvre d'habiletés plus générales dont le défaut entraîne des échecs répétés. Les perspectives cognitivistes

et néo-structuralistes mettent l'accent sur l'empan de la mémoire de travail et sur les processus d'attention et de contrôle. Gelman et Meck (1991, p.220) sont également sensibles à ces aspects du développement. Elles écrivent en effet que, malgré l'existence de compétences mathématiques bien définies, deux sources d'erreurs sont possibles : une compétence procédurale inadéquate au niveau de l'établissement des plans d'actions et une compétence interprétative limitée. Ces deux types de compétence sont à l'évidence impliquées dans le système modulaire de l'attention mentale de Pascual-Leone (cf. paragraphe 1.4.1.) où l'accent est mis sur les capacités conjointes d'activation des schèmes pertinents et d'inhibition de ceux qui sont «dangereux» ou pour le moins inadaptés. Ces compétences sont aussi impliquées dans les structures de contrôle exécutifs de Case (cf. 1.4.2.) qui régissent, compte tenu des connaissances intégrées en mémoire à long terme, les liaisons entre représentation du problème, objectif et stratégies de résolution - ces liaisons, à une étape donnée du développement, étant étroitement tributaires de la capacité de la mémoire de travail.

Outre l'évaluation des connaissances mathématiques, il semble bien que celle des capacités de mémorisation, d'attention et de contrôle soit nécessaire pour un examen approfondi des difficultés rencontrées par certains enfants.

4.4 Vers une évaluation diagnostique plus dynamique

L'expérience de Clements (1984), relatée au paragraphe 2.3., montre comment un apprentissage pas à pas permet à l'enfant «d'exprimer» ses capacités réelles. On a vu que K. Fischer (cf. 1.4.3.) appelle *niveau optimal de développement* la limite supérieure de complexité du répertoire des habiletés de l'enfant à une étape donnée. Ce niveau n'est exploité que lorsque le support contextuel est optimal. Les épreuves proposées au cours d'une évaluation classique sollicitent-elles ce niveau ? Une estimation qui impliquerait une collaboration constructive entre l'adulte-observateur et l'enfant ne rendrait-elle pas compte d'un état des lieux plus exact ?

La notion de zone proximale de développement de Vygotsky (1984, p.272-273) est intéressante à cet égard. Elle rend compte des rapports entre apprentissage et développement au cours de l'histoire sociale de l'enfant. Celui-ci peut, à tout moment, être caractérisé sous deux aspects : 1/ son niveau actuel, fonctionnel, tel qu'il peut être évalué à l'aide d'épreuves standardisées ou non ; 2/ son niveau de développement potentiel correspond *à ce qu'il est capable de faire avec un adulte à un certain moment* et sera capable de réaliser seul ensuite. Cette potentialité, plus ou moins actualisable selon l'interaction, que Vygotsky nomme *zone proximale de développement* correspond au *niveau optimal* de K. Fischer. A noter que K. Fischer invoque un contexte optimal d'actualisation et non une intervention de l'adulte.

On peut se demander si une interaction préparée et programmée ne permettrait pas, au cours de l'examen, l'estimation plus exacte des connais-

sances déclaratives et procédurales et *des points d'achoppement*. Elle serait à conduire au sein ou en complément d'une évaluation diagnostique classique. Il me semble aussi que le contenu, lorsqu'il est déterminé dans un domaine précis, des concepts de zone proximale, niveau optimal et niveau fonctionnel, serre de plus près la réalité du développement que l'opposition très générale compétence/performance.



Apport et limites de la psychologie cognitive à l'analyse des procédures de calcul et de résolution de problème par les élèves

Jean Retschitzki

Introduction

Il est devenu banal de souligner l'interdépendance entre l'évaluation et l'enseignement; l'évaluation joue un rôle considérable dans le processus éducatif, que ce soit directement pour l'orientation et la sélection des élèves, ou indirectement par les messages implicites que véhicule le choix de tels ou tels contenus lors des examens, pour ne citer qu'un exemple parmi beaucoup d'autres. S'interroger sur l'évaluation diagnostique des apprentissages, c'est déjà faire un pas de plus et privilégier une approche centrée sur l'individu en tant qu'apprenant et se fixer comme objectif de mettre en évidence les acquisitions résultant du processus d'apprentissage. Si les travaux psychologiques classiques n'ont guère permis de traiter de manière satisfaisante ce type de question, les développements qu'a connus la psychologie cognitive depuis près de trente ans semblaient davantage porteurs de promesses quant à la mise à disposition des praticiens de l'éducation d'outils susceptibles de guider leur action. Dans les quelques pages qui suivent nous tenterons de faire un bilan des contributions de la psychologie cognitive à ce domaine particulier d'application.

A propos des changements intervenus dans différents domaines des sciences humaines au cours des 40 dernières années, il est devenu courant de parler de «révolution» cognitive (Gardner, 1985). Nous manquons de recul pour en apprécier pleinement l'ampleur et juger si Varela a raison en pensant qu'il s'agit de *«la plus importante révolution conceptuelle et technologique depuis l'avènement de la physique atomique, ayant un impact à long terme à tous les niveaux de la société»* (Varela, 1989, p. 21).

Lorsqu'on s'interroge sur les apports de la psychologie cognitive, il convient en premier lieu de préciser de quelle psychologie cognitive il est question. En effet, au sens large la psychologie cognitive englobe un ensemble très vaste de domaines et de travaux dont il est bien difficile d'apprécier en général les apports et les limites, vu la diversité qui les caractérise. Nous adoptons plutôt une terminologie consistant à distinguer, à la suite de Wagner et Sternberg (1984), trois conceptions de l'intelligence ayant des implications pour l'enseignement : la perspective psychométrique, la perspective piagétienne et la perspective cognitiviste au sens strict souvent désignée par l'expression «traitement de l'information». Cette dernière approche est, depuis les années 60, l'une des tendances dominantes en psychologie, ayant supplanté l'approche comportementaliste. Cet «impérialisme» a même touché des domaines aussi divers que les pratiques thérapeutiques (Beck, 1976; Ellis, 1962) ou l'explication des troubles du langage (Ellis, 1989). On peut mentionner que cette domination d'une approche cognitiviste n'est pas spécifique à la psychologie, mais qu'on l'observe également dans plusieurs autres disciplines.

Pour être complet, on se doit de mentionner le fait que l'apogée du courant cognitiviste semble dépassée au début des années 90 et que des critiques se sont élevées contre différents aspects de cette approche, de nouvelles

tentatives étant entreprises dans différentes directions pour dépasser les difficultés auxquelles elle se heurte.

Il sera question ici plus spécifiquement des apports de la psychologie cognitive au sens strict du courant «traitement de l'information» dans un domaine particulier, à savoir l'analyse des procédures des élèves. Cette discussion semble d'autant plus légitime que ce courant est marqué par un souci d'application plus marqué que dans le cadre des approches qui peuvent lui être comparées.

1. SPÉCIFICITÉS ET APPORTS DE L'APPROCHE COGNITIVE

L'un des représentants les plus typiques du courant cognitiviste, John Anderson (1980, 1983) a souligné la nécessité d'une analyse abstraite, seule à même de rendre compte des conduites d'un sujet au niveau qui nous intéresse. Pour expliquer les conduites cognitives on pourrait se tourner vers les mécanismes physiologiques sous-jacents. Non seulement on se heurterait à des difficultés techniques sérieuses, mais il s'agirait aussi d'un niveau d'analyse trop détaillé pour être utile. Un programmeur qui donne trop de détails sur le déroulement de son programme ne nous apprend pas l'essentiel; il convient plutôt de rester à un niveau suffisant d'abstraction. Les psychologues cognitivistes sont donc à la recherche du bon ensemble de concepts susceptibles de décrire et d'expliquer l'intelligence humaine. La physiologie peut poser des contraintes que devront satisfaire les théories mais non déterminer les théories elles-mêmes (p. ex. la quantité d'information que le cerveau peut contenir est susceptible d'exclure certaines théories, mais non de suggérer la bonne théorie).

La tâche des psychologues cognitifs est de nature inférentielle. A partir de l'observation des conduites, il s'agit de décrire les mécanismes sous-jacents. En psychologie cognitive, une théorie est un système conçu pour résumer et systématiser divers faits connus et faire des prédictions au sujet de phénomènes à découvrir. Une théorie est basée sur les faits connus, mais plusieurs théories peuvent rendre compte des faits. Cela est particulièrement vrai en psychologie cognitive. Le psychologue propose des théories sur ce qui se passe dans la tête du sujet, sur la base du comportement externe. Il n'y a aucun moyen de savoir si une théorie cognitive décrit vraiment ce qui se passe à l'intérieur.

Toujours selon Anderson, une théorie cognitive diffère donc de la vérité de deux manières : (1). Même en cas de prédiction correcte, il n'y a pas de certitude de l'adéquation de la théorie avec la réalité des processus internes. Beaucoup de mécanismes différents peuvent en effet produire le même comportement extérieur (ceci peut être démontré dans le cadre de la théorie des automates). (2). Même si une théorie rend bien compte des données, il

existe toujours une chance qu'elle n'arrivera pas à prédire correctement le résultat d'une nouvelle expérience et devra donc être révisée.

L'approche cognitive a eu un impact très important dans pratiquement tous les domaines de la psychologie. Il est sans doute trop tôt pour faire un bilan définitif des apports de cette nouvelle perspective d'autant plus que son développement est loin d'être achevé. On peut néanmoins se risquer à évoquer quelques éléments qui semblent incontestables, d'une part en ce qui concerne les méthodes et d'autre part au sujet de quelques résultats plus particulièrement marquants à notre point de vue.

Sur le plan méthodologique, l'approche cognitive en psychologie a apporté quelques innovations marquantes en tête desquelles il convient de mentionner d'une part la simulation des processus cognitifs d'un sujet confronté à une tâche donnée, et d'autre part le recours très fréquent à des modèles du fonctionnement mental. Il s'agit plus d'apports complémentaires que d'une révolution méthodologique. En effet, bien qu'on ait rarement recours en psychologie cognitive aux analyses statistiques complexes, cela ne signifie pourtant pas que l'expérimentation ait été abandonnée. Mais diverses techniques d'analyses ont été soit créées, soit utilisées préférentiellement.

En fonction de la centration sur les processus, plutôt que les structures, on a eu plus particulièrement recours à des variables comme les temps de réaction ou les mouvements oculaires. De même les protocoles verbaux ont connu un regain d'intérêt de la part des chercheurs, notamment ceux travaillant sur la résolution de problèmes, tout comme l'analyse qualitative des données (types d'erreurs, etc.). Comme autre tendance caractérisant la plupart des travaux cognitivistes, on peut noter une centration sur les comportements individuels envisagés dans leur complexité, plutôt que l'intégration d'un grand nombre de données provenant d'échantillons de sujets.

2. QUELQUES EXEMPLES D'ANALYSE DES PROCÉDURES

Pour mieux cerner l'apport de la psychologie cognitive dans le domaine qui nous intéresse ici, il semble indiqué d'examiner quelques exemples paradigmatiques de l'utilisation de l'approche cognitive pour l'évaluation des compétences et la description des procédures des élèves au prise avec des opérations ou des problèmes mathématiques.

Dans la suite de ce chapitre, nous avons choisi de développer et discuter brièvement quelques recherches parmi les plus représentatives à notre point de vue. Il va de soi que ce choix est arbitraire et donc critiquable. Un des critères qui a permis de les retenir comme représentatives des apports de la psychologie cognitive réside dans le fait qu'elles comportent des analyses d'une complexité telle qu'il aurait été difficile de les mener à bien sans le con-

cours de moyens informatiques. Les travaux retenus concernent d'une part les opérations arithmétiques qu'elles soient orales ou écrites, et d'autre part les conduites observées lors d'épreuves inspirées peu ou prou de la théorie piagétienne.

2.1 Analyse chronométrique des opérations arithmétiques

Deux études consacrées à ce thème (Groen et Parkman, 1972; Woods, Resnick, et Groen, 1975) ont connu un succès d'estime considérable si l'on en juge par les nombreuses citations auxquelles elles ont donné lieu avant d'être répliquées et discutées. Dans les deux cas, le fonctionnement mental d'un groupe de sujets confrontés à une tâche arithmétique élémentaire (additions et soustraction de 2 nombres) est inféré à partir des temps de latence. Entre les solutions par comptage des tout petits et les solutions par simple évocation des résultats mémorisés qu'on peut observer dès que les tables sont suffisamment connues, il semble y avoir une période où les enfants calculent mentalement selon des procédures inventées par eux. Ce sont ces conduites que les deux groupes de chercheurs ont étudiées.

Ces deux groupes de chercheurs ont testé une famille de modèles possibles tant pour l'addition de deux nombres dont la somme est inférieure à 10 que pour la soustraction de deux nombres. Tous ces modèles postulent l'existence d'un compteur qui doit être initialisé à une certaine valeur, avant d'être incrémenté un certain nombre de fois, puis «lu» pour donner le résultat. En faisant l'hypothèse que les opérations élémentaires prennent un temps relativement constant, les différents modèles permettent de faire des prévisions sur l'allure des distributions des temps de latence.

La comparaison des configurations de temps de réaction observés chez différents groupes d'élèves avec ceux que prédisent les différents modèles a suggéré aux auteurs une interprétation selon laquelle les enfants les plus jeunes (1ère année) utilisent déjà le «MIN MODEL» pour l'addition, qui se caractérise par le fait que les élèves déterminent d'abord quel est l'addende maximum, ce qui leur permet ensuite de n'incrémenter qu'un nombre minimum de fois. Les seules exceptions concernent les additions doubles (2+2 / 3+3 / 4+4) qui prennent nettement moins de temps. Cette exception serait explicable par le fait que les sujets parviennent à mémoriser les résultats de ces opérations plus rapidement que ceux des autres calculs; ils utiliseraient donc un processus de nature différente. Ce même modèle conviendrait également pour les résultats d'additions jusqu'à des sommes de 18 et avec des enfants de 4 1/2 à 9-10 ans, comme d'autres recherches l'ont indiqué ultérieurement.

Siegler a consacré différents articles à la critique de l'usage exclusif de la chronométrie mentale pour inférer les procédures des sujets (Siegler, 1987; 1989). Ayant répliqué l'expérience en question, il a en outre interrogé les sujets immédiatement après la réponse pour connaître la stratégie qu'ils avaient utilisée. Les résultats de ses études montrent que les sujets, loin de suivre pra-

tiquement tous la même procédure comme le suggéraient les résultats des études originales, utilisent une grande diversité de stratégies. Siegler montre ainsi les dangers d'une analyse superficielle. Utilisée de manière exclusive, la technique chronométrique peut conduire à des interprétations erronées, alors que si elle est utilisée conjointement à d'autres techniques, elle peut se révéler féconde.

Ces résultats de Siegler confirment d'ailleurs une étude plus ancienne de Lankford (1972) sur les quatre opérations élémentaires auprès de 176 élèves de 7ème année qui montrait déjà que les stratégies de calcul sont très individuelles et ne suivent pas forcément les modèles canoniques appris en classe ou dans les manuels. En outre, ces élèves utilisaient fréquemment le comptage dans les calculs et Lankford avait observé des différences entre les stratégies des bons calculateurs et celle des moins bons.

Dans le même ordre d'idées, on peut mentionner les travaux ayant trait à ce qu'on appelle l'étude des savoirs quotidiens. A la suite de recherches entreprises initialement dans le domaine de la psychologie interculturelle, diverses équipes se sont penchées sur la comparaison des connaissances scolaires et des connaissances acquises d'une manière informelle à travers différents types d'activités (petit commerce sur les marchés, tissage, poterie, etc.). En restant proche du domaine des connaissances mathématiques, on peut citer notamment les travaux du groupe de Recife (Carraher, Carraher & Schliemann, 1987; Nunes, Schliemann & Carraher, 1993). Ces chercheurs ont comparé les performances d'enfants des écoles de l'état de Pernambuco dans le Nordeste brésilien âgés de 8 à 13 ans pour divers problèmes d'arithmétique. Ces travaux ont mis en évidence des procédures bien différentes selon le contexte de présentation des opérations et la supériorité des procédures orales inventées par les enfants pour résoudre les problèmes par rapport aux algorithmes écrits enseignés en classe. D'autres travaux ont confirmé la supériorité des connaissances quotidiennes sur les connaissances purement scolaires.

On peut voir dans ce type de travaux l'origine lointaine du courant actuel qui, s'inspirant notamment des conceptions de Vygotsky, met l'accent sur l'importance du contexte dans l'acquisition des connaissances (situated learning, situated cognition) et sur la nature nécessairement sociale de la connaissance. Bien que plusieurs des auteurs de ces travaux récents aient appartenu au courant cognitiviste (en particulier Brown et Collins), ils ont opéré une nouvelle révolution et la filiation avec leurs anciennes recherches est parfois difficile à retracer.

2.2 Modèles de diagnostic de Brown et Burton

A la fin des années 70, l'article de Brown et Burton (1978) sur les modèles de diagnostic a connu un grand succès dans les milieux intéressés. A partir de l'idée que la compréhension de la nature des erreurs des élèves est une condition *sine qua non* pour une remédiation efficace, Brown et Burton

se sont efforcés de parvenir à un diagnostic automatique des erreurs d'un élève sur la base de l'analyse de l'ensemble des réponses obtenues à un test comprenant 15 soustractions écrites. Ils ont élaboré un programme (BUGGY) destiné à analyser les patterns de réponses des élèves pour découvrir les «bugs» qui se sont infiltrés dans la procédure algorithmique.

Ils font observer que si additionner ou soustraire deux nombres n'est pas très difficile, diagnostiquer à coup sûr les erreurs dans ces opérations peut se révéler assez subtil. Contrairement à l'opinion répandue chez de nombreux parents et enseignants selon laquelle les erreurs proviennent du fait que les élèves ne suivent pas très bien les procédures apprises, ces auteurs font en outre l'hypothèse que les élèves suivent au contraire très bien les procédures, mais qu'il s'agit souvent de procédures (au moins en partie) fausses.

Leur but est de parvenir, pour chaque élève considéré, à un modèle de diagnostic, c'est-à-dire une représentation de la connaissance procédurale qui décrit l'intériorisation de la procédure par l'élève comme une variante d'une version correcte de cette procédure. Ayant analysé les résultats de près de 1300 élèves au moyen d'une technique sophistiquée reposant sur l'intelligence artificielle, ils ont pu montrer que l'on pouvait déceler des erreurs systématiques susceptibles de donner lieu à une remédiation spécifique chez près de 40% des élèves.

Selon le point de vue que l'on adopte, ce pourcentage peut sembler faible puisque pour la majorité des élèves on n'a pas de diagnostic, ou au contraire on peut être impressionné par le fait qu'il soit possible de mettre en évidence des déviations systématiques de procédures chez des élèves dont on ne connaît que leurs réponses à 15 opérations. Les données analysées provenaient en effet d'une étude menée antérieurement au Nicaragua par un autre groupe de recherche et dont les résultats bruts avaient été stockés sur cartes perforées. Donc les auteurs de l'analyse ne pouvaient rien connaître des élèves en question et si les diagnostics avaient pu être validés et se révéler corrects, un tel pourcentage aurait permis d'entreprendre auprès de nombreux élèves des actions de remédiation relativement efficaces car bien ciblées.

Brown et Burton ont également pu identifier les types d'erreurs les plus fréquents ce qui pourrait se révéler utile aux enseignants soucieux de prévenir de telles conséquences par des exercices appropriés. Il me semble qu'il était également important, du point de vue pédagogique, d'attirer l'attention sur le fait que des tests moyennement réussis (permettant souvent d'obtenir des notes suffisantes !) peuvent masquer des acquisitions partiellement défectueuses. Même si des réflexions de ce type étaient possibles avant cette étude, les auteurs me semblent en avoir fait une démonstration particulièrement convaincante.

Sans doute plus intéressés par les questions fondamentales, ils se sont ensuite attachés à simuler la genèse des erreurs systématiques chez le sujet (Brown & VanLehn, 1980, 1982; VanLehn, 1987), s'écartant ainsi des aspects pédagogiques et diagnostiques pourtant très présents dans le premier article.

En effet, si les points rappelés ci-dessus sont relativement bien connus, l'aspect plus particulièrement pédagogique de l'article a lui presque totalement passé inaperçu. Il comporte pourtant des idées intéressantes qui n'ont pas été totalement exploitées à notre connaissance. Parmi celles-ci, je mentionnerai en particulier les conseils pour la mise au point de tests apportant des informations riches et interprétables de manière univoque (Brown & Burton, 1978, pp 183-184), ainsi que l'usage du programme à des fins de formation des futurs enseignants. L'augmentation de la puissance des ordinateurs personnels à laquelle nous avons assisté ces dernières années, et qui se poursuit voire s'accentue, autoriserait probablement la vulgarisation et une large diffusion de tels produits informatiques auprès des milieux de l'enseignement, ce qui eut été impensable il y a une quinzaine d'années lors du développement des prototypes de ces programmes.

Une approche très similaire à celle de Brown et Burton a été adoptée par Hasemann (1986) dans son article sur les erreurs dans le calcul des fractions; toutefois cet auteur s'est inspiré pour ses analyses de concepts tels que celui de «frame» développé par Minsky (1975) et de «mécanismes hypothétiques» de Davis et McKnight (1979). Il a également trouvé une grande diversité dans les conduites des enfants qu'il attribue aux représentations des concepts mathématiques propres aux différents élèves.

Dans son ouvrage sur l'évaluation formative des apprentissages, Scallon (1988) fait une revue de littérature très complète des travaux consacrés à l'analyse des erreurs systématiques. Une telle démarche a été adoptée pour traiter de très nombeux domaines appartenant soit aux mathématiques, soit à l'apprentissage de la langue. Scallon souligne en conclusion du chapitre consacré à cet examen combien cette approche est intéressante à titre d'esprit devant guider la réflexion des enseignants plutôt que comme technique d'évaluation prête à l'usage.

2.3 Reconstitution de règles ou stratégies suivies par les sujets

De nombreuses recherches se sont centrées sur les stratégies de résolution de problème et il serait illusoire de tenter ici d'en faire l'inventaire. Notre choix s'est porté sur une étude de Siegler dont la méthode est particulièrement intéressante pour notre propos. Reprenant le problème de l'équilibre de la balance étudié par Inhelder et Piaget (1955), avec une méthodologie profondément renouvelée, Siegler (1988) s'est efforcé de reconstituer les «programmes» ou règles que semblent suivre les sujets pour produire leurs réponses à une tâche où il s'agit de formuler des prévisions quant à l'état du dispositif en fonction de diverses répartitions des poids.

Siegler a élaboré quatre modèles de complexité croissante sous la forme d'un arbre de décision que le sujet est censé parcourir en répondant à des questions élémentaires sur le dispositif (Les poids de part et d'autre de la ba-

lance sont-ils égaux ? / Les distances sont-elles égales ?). En fonction des réponses apportées à ces questions, le sujet parvient à fournir une réponse à la question de l'expérimentateur. Chacun des quatre modèles postulés par Siegler correspondait en gros aux étapes du développement décrites par les auteurs genevois.

Ces modèles ont la particularité de permettre à l'auteur une prédiction de l'allure des courbes de développement avec l'âge différenciées selon la nature des situations présentées aux sujets. C'est ainsi que pour l'une des catégories de problème (« conflit-poids »), les hypothèses basées sur une analyse de type traitement de l'information prévoyaient une régression avec le modèle III. Les résultats confirment cette évolution, bien que le redressement final n'ait pas été observé, même avec le groupe des sujets les plus âgés (15-16 ans).

En appliquant des critères assez stricts, Siegler a montré que les sujets pouvaient dans leur grande majorité être «diagnostiqués» comme utilisateur d'un des modèles. L'absence quasi totale du modèle optimal même chez les plus âgés expliquerait le fait que la situation «conflit-poids» ne donne pas lieu à une réussite majoritaire après la baisse de performance en fonction de l'âge.

Une équipe québécoise a répliqué cette étude de Siegler en critiquant certains aspects de son approche (Larivée, Normandeau, Roulin & Longeot, 1987). Selon eux, le modèle n'intègre pas toutes les règles possibles et il passe sous silence la variabilité intra-individuelle. A travers leur analyse, les auteurs mettent en évidence trois autres règles dont celle de l'additivité, consistant à «additionner le poids et la distance de chaque côté du fléau et à conclure que la balance penchera du côté où la somme est la plus élevée» (Larivée & al, 1987, pp. 521-522). Ils soulignent aussi la trop grande généralité de la Règle III du modèle de Siegler puisque parmi les 77 sujets qui seraient classés comme utilisateurs de la Règle III avec l'analyse de Siegler, 33 sujets utilisent en fait la règle de l'additivité.

Des tentatives ont été entreprises pour appliquer un schéma similaire au cas de l'utilisation d'un traitement de texte (Desmarais, Larochelle, & Giroux, 1987). Par la simple observation (automatique) des actions d'un utilisateur, il s'agit de déterminer l'état actuel de ses connaissances, des facilités offertes par le programme, afin de pouvoir lui donner quelques conseils judicieux. Le but est de parvenir à programmer un «coach» simulant l'activité d'un moniteur qui chercherait à faire progresser les utilisateurs en leur suggérant des méthodes plus économiques ou des raccourcis, etc.

2.4 Étude des stratégies des joueurs d'awélé

Au-delà du problème de la balance, cette méthode, «rule-assessment» ou «par élaboration de règles» selon la traduction adoptée par Larivée, est susceptible de s'appliquer à d'autres types de tâches. Nous illus-

trerons une telle transposition par la présentation un peu plus détaillée d'une recherche menée en Côte d'Ivoire sur les stratégies des joueurs d'awélé, à l'occasion de laquelle nous avons également eu recours à une technique similaire en complément d'autres approches pour tenter d'identifier les stratégies des enfants dans la pratique de ce jeu qui nécessite notamment des raisonnements de type numérique (Retschitzki, 1990). En raison de la pauvreté des verbalisations des joueurs quant à leurs stratégies, il s'est avéré très précieux de disposer d'un moyen supplémentaire d'investigation à ce sujet.

2.4.1 *Présentation du jeu d'awélé*

Le jeu d'awélé appartient à la classe des jeux de semailles. Nous avons étudié la variante qui se joue chez les Baoulés de Côte d'Ivoire. L'awélé se joue à 2, avec un «tablier» (une planche de bois creusée de 2 rangées de 6 trous). Au départ chaque case contient 4 graines. A tour de rôle chacun des joueurs vide le contenu d'une case qui se trouve de son côté et sème les graines dans les cases suivantes, à raison d'une graine par case, dans le sens inverse des aiguilles d'une montre. Le but du jeu est de récolter plus de la moitié des graines chez l'adversaire. Une récolte est possible lorsque le joueur termine sa semaille dans une case qui contient déjà une ou deux graines portant ainsi le contenu à 2 ou 3 graines. Le joueur a de plus le droit de récolter le contenu de toutes les cases précédentes satisfaisant au même critère.

Pour appréhender les stratégies des joueurs, notamment des enfants, il était difficile d'avoir recours à la méthode «thinking aloud», souvent employée avec les sujets adultes dans les études cognitivistes sur des thèmes similaires. C'est pourquoi nous avons tenté de recueillir le maximum d'informations en combinant différentes techniques : observation directe, épreuve de connaissance du jeu, observation différée (grâce aux enregistrements vidéo), analyse détaillée des parties après transcription, interrogation de joueurs adultes, analyse de situations de jeu par les adultes, etc.

En plus, nous avons eu recours avec les enfants à une épreuve papier-crayon inspirée des problèmes d'échecs qu'on trouve dans les journaux ou périodiques. Ce sont les données de cette épreuve qui seront discutées plus en détail ci-dessous.

2.4.2 *Méthode*

Nous nous sommes efforcés de constituer un ensemble de situations aussi varié que possible pour déceler les règles que les sujets semblent suivre en choisissant un coup dans différentes circonstances du jeu. On présentait aux sujets un ensemble de configurations de jeu représentées symboliquement. Les sujets devaient simplement indiquer par une flèche quelle case il leur semblait préférable de jouer du point de vue du joueur dont les quantités de graines étaient représentées dans la rangée inférieure.

La confrontation des réponses des sujets avec l'analyse des sujets devrait permettre de formuler des hypothèses sur le type de processus cognitif utilisé par les sujets. Dans quelle mesure s'en tiennent-ils à une analyse statique ou au contraire sont-ils capables d'anticipation, et si oui, jusqu'à quelle profondeur ? Tel est le genre de question que le dépouillement de cette épreuve devrait aider à clarifier.

Ce qui caractérise cette épreuve et rend son exploitation relativement complexe, c'est qu'il n'y a pas à proprement parler de «bonne réponse», mais que l'évaluation d'une réponse dépend de l'objectif que s'assigne le joueur.

L'une des difficultés de cette démarche réside dans le fait qu'on ne peut pas interpréter à coup sûr une réponse isolée. Le choix du même coup peut en effet s'expliquer de différentes manières, surtout dans un jeu où le nombre de coups possibles est au maximum de 6. Ce n'est que par la prise en compte de la totalité des réponses qu'on peut espérer approcher le «programme» réel des sujets, pour autant que leur manière d'opérer le choix du meilleur coup soit relativement cohérente et stable tout au long de l'épreuve.

Confronté à une configuration qui représente une situation du jeu d'awélé, les sujets doivent choisir au maximum parmi 6 réponses possibles. Pour établir les différents modèles qui doivent être confrontés aux réponses réelles, on peut formuler des hypothèses sur les informations traitées par les sujets, l'ordre dans lequel elles sont prises en considération.

Compte tenu des règles du jeu, un certain nombre de critères semblent pertinents :

- Priorité à l'aspect offensif (aspect défensif négligé) ou l'inverse.
- Décision dès qu'un coup apparaît productif vs décision après examen global.
- Comparaison des aspects offensifs et défensifs.
- Examen des cases de gauche à droite ou l'inverse.
- Perception de toutes les menaces potentielles ou ignorance des menaces de moyenne ou longue distance

2.4.3 *Modélisation des configurations de réponses*

Dans le but de trouver les règles que semblent suivre les sujets, des modèles hypothétiques ont été élaborés et les prédictions qu'ils fournissent confrontées avec les réponses réelles des sujets. Le Tableau 1 présente quatre exemples de modèles pour illustrer le principe de la démarche :

TABLEAU 1

Exemple de modèles des stratégies pour l'épreuve des problèmes d'awélé

M01	Les cases sont examinées dans l'ordre droite-gauche Chaque case est testée pour les récoltes possibles Dès qu'une récolte possible est décelée, on choisit cette case Si aucune récolte n'est décelée, pas de prévision

M14	Tester chaque case pour les menaces selon l'ordre G-D Dès qu'une perte est possible, jouer la case menacée Tester chaque case pour les récoltes selon l'ordre D-G Dès qu'un gain est possible, jouer la case Sinon, pas de prévision

M26	Faire l'inventaire des menaces et des récoltes Comparer gain maximum et perte maximum Jouer la case avec conséquence maximale En cas d'égalité des effets, priorité à la récolte Si plusieurs cases ont le même effet, choisir selon l'ordre G-D Sinon, pas de prévision

M42	Appliquer l'algorithme minimax avec une profondeur de 2 Choisir le meilleur coup selon cette procédure En cas d'égalité, choisir la case la plus à droite

Un exemple de problème n'est pas inutile pour concrétiser la démarche suivie. Le problème No 126 se présentait de la manière suivante (cf. Figure 1).

1	**4**	**6**	**6**	**2**	**0**
2	**1**	**1**	**2**	**1**	**6**

FIGURE 1

La situation du problème 126

Dans le cas de ce problème, les 4 modèles prédisent les réponses suivantes, ainsi que l'indique le Tableau 2.

TABLEAU 2

Réponses prédites pour le problème 126 par les quatre modèles retenus

M01	M14	M26	M42
6	1	4	3

2.4.4 *Principaux résultats*

Pour avoir une vision synthétique des tendances qui se dégagent de l'analyse, les différents modèles peuvent être regroupés en trois grands groupes selon l'approche du problème qui les caractérise :

1. Les modèles caractérisés par un examen partiel de la situation
2. Les modèles caractérisés par un examen global de la situation
3. Les modèles dérivés d'une analyse exhaustive (algorithme minimax)

En analysant les concordances entre les prédictions basées sur les différents modèles et les réponses observées, on peut dégager les principales tendances des résultats. Pour chaque sujet, on calcule autant de scores de concordance qu'il y a de modèles. Une correspondance complète entre les prédictions des modèles et les réponses réelles est relativement peu fréquente. On ne trouve un ou des modèles expliquant la quasi totalité des réponses que chez 6 sujets sur 28.

TABLEAU 3

Répartition des sujets selon la catégorie de modèle

Aspect privilégié	Examen de la situation					
	Partiel			Global		
Offensif	1112	1313			1305	
	1113	1315				1514
	1114		1502			
Défensif	1105			1111		
	1115			1303		
		1302				
Offensif + défensif				1101	1304	1501
				1102	1312	1504
				1103	1314	1511
				1104		1512
Sujets inclassables		1311	1530	1513		

Le Tableau 3 indique la répartition des sujets dans les grandes catégories de modèles. On peut relever que parmi les 9 sujets qui semblent choisir leur coup après un examen superficiel on trouve surtout des sujets de 11 ans. Inversement parmi ceux qui semblent procéder à un examen global et en tenant compte aussi bien des aspects offensifs que des aspects défensifs, on trouve la majorité des sujets de 15 ans.

On notera également que parmi les sujets inclassables figure le sujet 1503, le meilleur joueur de l'échantillon. Cela pourrait indiquer qu'aucun modèle ne correspond au meilleur choix possible. Il pourrait s'avérer nécessaire d'élaborer de nouveaux modèles pour simuler les comportements non encore expliqués.

En examinant les résultats selon les modèles et la fréquence avec laquelle les différents sujets semblent raisonner d'une telle manière, on peut remarquer la prédominance des modèles accordant la priorité à l'aspect offensif, ainsi que la grande fréquence des modèles reposant sur une analyse partielle de la situation proposée.

Inversement, les modèles reposant sur une anticipation exhaustive de type minimax sont presque totalement absents. On ne trouve de bons scores que chez quelques sujets pour une feuille donnée. Cela semble indiquer que ce n'est pas en se livrant à une étude de tous les coups possibles que les sujets se sont déterminés. Ou pas seulement de cette manière, mais que d'autres éléments ont été pris en considération, ce qui n'est psychologiquement guère surprenant.

2.4.5 *Discussion*

A l'issue de l'analyse esquissée ci-dessus, quelques remarques s'imposent quant à l'utilité de la méthodologie retenue. En premier lieu, on peut souligner la richesse des indications fournies par le dépouillement de cette épreuve, alors que l'effort demandé aux sujets s'est révélé relativement peu important et que la passation a duré très peu de temps. La principale limite de cette épreuve pourrait résider dans l'absence de représentativité des réponses par rapport à la conduite qui nous intéresse, à savoir le choix d'un coup dans une situation réelle. La situation d'examen est en effet assez éloignée de la situation d'une vraie partie d'awélé. Outre les différences objectives évidentes (absence d'adversaire, test papier-crayon, absence d'enjeu, etc.), une différence très importante réside dans l'absence d'un contexte. Dans une partie réelle, le joueur connaît le score approximatif de chacun des adversaires et ses choix sont très probablement influencés par cette connaissance. Selon qu'il mène largement, qu'il est distancé ou que les positions sont équilibrées, il pourra décider ou non de jouer tel coup dont l'issue lui semble incertaine. Dans la situation des problèmes d'awélé, rien de tel ne vient orienter les choix des sujets. Incontestablement, il y a là un aspect un peu artificiel de la situation. On peut cependant penser que cette difficulté est peut-être moins importante dans le cas des garçons de 9 à 15 ans que ce ne serait le cas pour des

adultes, sans doute plus sensibles à la situation stratégique générale et dominant mieux l'ensemble des aspects pertinents d'une situation de jeu. De plus, les résultats ont bien montré que les sujets ne répondaient pas de manière aléatoire.

Différentes mesures pourraient être envisagées pour améliorer la correspondance entre prédictions et réponses; il conviendrait en premier lieu d'affiner les modèles, d'en créer de nouveaux; on pourrait aussi tenter d'élaborer un ensemble de problèmes conçus de manière plus systématique en vue de mieux vérifier des aspects particuliers des conduites visées. Mais cette première expérience était sans doute nécessaire pour vérifier qu'un tel effort mérite d'être entrepris.

Au total, à condition d'utiliser ce type d'épreuve comme complément à d'autres approches des stratégies des joueurs, elle nous semble être très utile et transposable à d'autres situations ou jeux et dans des contextes variés. Mais il convient d'insister sur le fait que cette méthode nécessite une analyse approfondie du domaine étudié et peut donc se révéler assez coûteuse.

3. ÉVALUATION DE CES TRAVAUX

Au terme de cette présentation de quelques recherches classiques représentatives de l'approche cognitive, il convient de tenter d'établir un bilan (au moins provisoire) des forces et des faiblesses de cette approche pour l'évaluation des procédures. Précisons que ces remarques sont avant tout valables pour les retombées directes des travaux cognitivistes, sans prétendre porter un jugement au sujet de l'influence que la psychologie cognitive a pu avoir, a et aura encore sur les très nombreuses recherches des spécialistes de l'évaluation pédagogique.

Les efforts entrepris dans le cadre du courant cognitiviste pour expliquer de manière plus fine les conduites des sujets dans différentes situations ont permis une meilleure connaissance des manières de fonctionner des sujets. Les conduites apparaissent plus complexes que les descriptions antérieures ou les études de type psychométriques ne le laissaient supposer. Cela explique le succès mérité de certains des travaux mentionnés précédemment lors de leur parution; ils donnaient véritablement l'impression d'une percée significative dans la connaissance psychologique.

C'est particulièrement vrai dans le cas de la recherche de Brown et Burton, qui avait permis de mettre en évidence des particularités très riches de résultats obtenus bien auparavant et ce sans aucun contact direct avec les élèves concernés. Le «miracle» aurait paru plus convaincant encore si l'on avait pu procéder à une validation des résultats de l'analyse et si une remédiation avait pu être entreprise auprès des élèves concernés. Mais les données avaient été mises à disposition par un autre groupe de recherche et la collecte des données était trop ancienne pour permettre une telle démarche.

Même s'il ne s'agit pas d'une découverte entièrement nouvelle, l'accent mis par le courant cognitiviste sur les aspects procéduraux a fortement contribué à enrichir notre connaissance dans le domaine des procédures de résolution de problème. C'est ainsi que la «créativité» dont font preuve les élèves qui inventent des procédures pour se débrouiller dans des situations limites est intéressante à observer. Les procédures enseignées à l'école ne sont que rarement suivies scrupuleusement, en tout cas pas immédiatement. Les élèves doivent d'abord se les approprier. D'autre part leurs performances dans des tâches scolaires ne correspondent pas toujours à leurs compétences réelles comme l'ont montré les observations dans le milieu extra-scolaire.

Enfin il faut créditer les études cognitives du fait d'avoir posé les problèmes d'une manière renouvelée et permis ainsi d'entreprendre des recherches d'une ampleur nouvelle dans des domaines autrefois trop complexes pour être étudiés scientifiquement.

Néanmoins ces travaux ont montré des limites qu'il convient également de mentionner et de discuter. On a vu les dangers, bien soulignés par Siegler (1987; 1989), d'une approche basée exclusivement sur une méthode chronométrique. Cela souligne la nature des théories en psychologie cognitive dont l'adéquation avec la réalité des processus sous-jacents reste toujours à confirmer. Cette remarque ne concerne bien entendu pas que les opérations arithmétiques et devrait nous inciter à la plus grande prudence et à toujours tenter de recouper les indications d'une approche par celles d'une autre. Je plaiderai donc pour une approche pluri-méthodologique.

Malgré les points positifs incontestables des travaux rendus possibles par les nouvelles technologies et les paradigmes de nature cognitiviste, nous ne disposons pas réellement d'outils permettant à chaque enseignant d'améliorer son activité d'accompagnement des élèves confrontés à des difficultés de maîtrise des opérations ou des situations de résolution de problème. Les travaux comme ceux de Brown et Burton comportent pourtant des indications sur la possibilité de construction d'épreuves de contrôle réellement informantes non seulement sur le classement des élèves mais encore sur leurs difficultés spécifiques. Il est regrettable que ces pistes n'aient pas été explorées de manière suffisante à notre connaissance. Les études en sont généralement restées au niveau de prototypes intéressants surtout du point de vue des recherches fondamentales.

Dans un récent article, Bejar et Braun (1994) indiquent que des synergies leur semblent possibles entre l'évaluation et l'enseignement en raison de l'intersection des besoins entre ces deux types d'activités (évaluation des performances, feed-backs aux apprenants, etc.). Mais il semble encore s'agir plus d'une croyance et d'un voeu que du constat de réalisations effectives.

On ressent donc une impression d'inachevé. Le versant application pédagogique des études semble surtout être mis en avant dans l'introduction et la conclusion des travaux, comme pour les légitimer ou pour justifier l'obtention des crédits de recherche. Mais la plupart du temps, il manque un réel

suivi des tentatives d'application pratique. Une des raisons pouvant expliquer le peu d'impact des recherches de psychologie cognitive fondamentale sur les travaux dans le domaine de l'évaluation réside peut-être dans la méconnaissance de ces travaux par les chercheurs en place dans les instituts ou services de recherche pédagogique; si c'est le cas, avec le temps on devrait constater des évolutions significatives, car la psychologie cognitive a maintenant largement pénétré des cercles de plus en plus étendus. Il y a donc lieu d'être relativement optimiste quant à la réalisation prochaine d'applications répondant aux attentes maintes fois exprimées à cet égard.

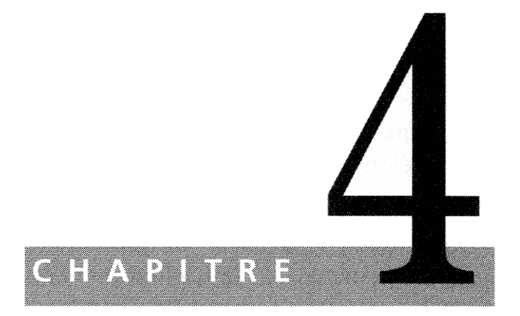
4. TENDANCES RÉCENTES

Au début des années 90, il apparaît toutefois que l'approche cognitive est à son tour remise en question. Neisser, dans un article récent, résume ce fait en écrivant «*Cognitive psychology is in a period of rapid and confusing change. Our old models... are being challenged by a flood of new discoveries.*» (Neisser, 1994, p. 225). Il est bien difficile de savoir si l'approche cognitiviste connaît une simple crise de croissance ou si elle est sur la pente d'un réel déclin. Parmi les évolutions récentes s'écartant notablement de l'orthodoxie cognitive, il convient de mentionner le développement des théories néo-piagétiennes, l'importance croissante des modèles néo-connexionnistes et l'apparition déjà citée plus haut du courant de «situated cognition».

La conception connue sous le nom de «distributed cognition» (Salomon, 1993) mérite une mention particulière car si on la prend au sérieux, le problème de l'évaluation en est singulièrement compliqué. En effet dans son expression extrême, elle revient à nier qu'il soit pertinent de considérer la connaissance d'un individu, puisque toute connaissance est nécessairement distribuée entre plusieurs cerveaux. Une telle approche pourrait conduire à renoncer à toute évaluation. A tout le moins il faudrait revoir fondamentalement les manières de procéder puisqu'il convient alors d'évaluer le système comme tel, et non les performances d'un individu isolé. Il y a là de jolis défis à relever pour les spécialistes de psychologie pédagogique.

Cela rejoint également les remarques de Papert dans son dernier ouvrage où il fait remarquer que les acquisitions importantes qui ont marqué sa vie intellectuelle se sont en fait déroulés en dehors de l'école (Papert, 1993, p. 23). Ce que nous tentons de mesurer, d'évaluer n'est-il de la même manière qu'un épiphénomène sans grande importance laissant de côté les aspects les plus essentiels des acquisitions cognitives ?

Nous manquons encore de recul pour juger valablement tant les apports du courant cognitif dont toutes les retombées ne sont pas encore connues que les apports potentiels des courants plus récents. Le bilan que nous avons tenté de tirer ne saurait donc être considéré que comme un instantané à réactualiser dans quelques années.



La mesure des Temps de Réponses aux calculs numériques élémentaires en situation scolaire

Jean-Paul Fischer

Introduction

L'introduction de l'outil informatique dans les écoles permet d'y reprendre une méthode d'investigation très fructueuse et très utilisée en psychologie cognitive : la mesure des Temps de Réponse (TR). Cette dernière s'applique particulièrement bien à l'étude de la mémorisation des faits numériques élémentaires, les tables si l'on préfère.

Ranschburg, au début de ce siècle, en 1903 à Budapest précisément, avait mesuré les TR sur des élèves de première année d'école, en difficulté (débiles) ou non (cf. Ranschburg, 1930). Il avait déjà observé la grande lenteur des premiers comparativement aux seconds et découvert quelques subtilités du traitement des additions comme le traitement particulièrement rapide et uniforme des doubles (e.g., 5+5). Mais les méthodes et problématiques piagétiennes ont ensuite, probablement, freiné cette utilisation des TR jusqu'à une époque récente où, consécutivement à l'article princeps de Groen et Parkman (1972) et grâce aux facilités matérielles nouvelles, on a pu assister à une explosion de la littérature (pour une revue, voir Ashcraft, 1992).

Toutefois, la plupart de ces recherches ont été menées dans une perspective davantage psychologique que pédagogique. Ceci sous-entend que, souvent, on n'interroge qu'une partie des élèves d'une classe; la passation est individuelle; on se limite à une seule opération arithmétique; on ne se préoccupe que rarement de ce qui est essentiel pour les pédagogues, à savoir le progrès des élèves.

Nous commencerons par présenter quelques problèmes de méthode et la manière dont nous les avons résolus. Ensuite, nous indiquerons deux applications immédiates de nos propositions : d'une part, le diagnostic pour une classe en début d'année, d'autre part la mesure des progrès tout au long de l'année voire de plusieurs années. Enfin, nous présenterons des applications plus générales de l'étude des TR, mais dont l'intérêt est néanmoins évident pour l'apprentissage, la différenciation pédagogique et la remédiation : ces applications concernent en effet le développement numérique, les différences individuelles et les sujets en difficulté.

1. QUESTIONS DE MÉTHODE

1.1 Tâche de vérification vs tâche de production

1.1.1 *Les contraintes matérielles d'une tâche de production*

La mesure des TR devient aujourd'hui, grâce à l'équipement informatique, praticable dans les classes. Néanmoins, un problème pratique important continue à se poser : il est difficile de demander aux élèves de produire la réponse, i.e. donner la réponse par exemple à 9+7. En effet, la mesure précise

des TR dans une tâche de production implique soit une réponse orale, soit des choix (de touches) multiples.

Dans le premier cas, il faut un dispositif spécial permettant l'arrêt du chrono au début de la réponse orale (voir une description dans Miller, Perlmutter & Keating, 1984, par exemple). Un tel dispostif risque de conduire à de nombreuses fausses alarmes (bruit ou réponse du voisin, éternuements, ...) et n'est guère utilisable qu'en conditions de laboratoire.

Dans le second cas, il se pose au moins deux problèmes :

1. Celui du temps d'accès aux différentes touches-réponses. Pour réduire ce temps au maximum, il faudrait un entraînement conséquent. Et même dans ce cas, on peut se demander s'il va être uniforme (i.e., indépendant de la touche sur laquelle on appuie) : on sait en effet que, même chez des secrétaires professionnelles, le temps de frappe (qui est beaucoup plus court que le temps de réaction réflexe) pour un clavier QWERTY varie de 190.5 à 236.2 ms si l'on considère les seules lettres de la première rangée, ou de 168.7 à 279.2 ms si l'on considère l'ensemble des lettres (Salthouse, 1984);
2. Celui de la taille du nombre-réponse. En effet, si ce dernier dépasse 9, i.e. si la réponse est un nombre à deux chiffres, le temps de recherche des touches contribuera pour une part extrêmement importante au TR final mesuré. Et si on se limite à des réponses inférieures à 10, les nombres utilisables pour les additions sont limités et, pire, les multiplications se trouvent presque exclues.

1.1.2 *Les spécificités d'une tâche de vérification*

La tâche de vérification - indiquer, par exemple, si 9+7 = 15 est juste - présente, comparativement à une tâche de production, certaines spécificités (Fischer, 1988b). La première, et probablement la plus importante, est qu'elle peut conduire à d'autres formes de vérification que le calcul procédural de la réponse ou sa récupération en mémoire déclarative. D'une part, parce que la fausseté d'une égalité peut être détectée sans faire le calcul, par exemple par un argument de parité (e.g., 6x7 = 43 est fausse car un produit dont l'un des termes est pair est nécessairement pair) ou par un jugement de plausibilité (e.g., 6x7 = 72 n'est pas plausible car 72 est supérieur à 10x7 = 70). D'autre part, parce que pour les opérations inverses - soustraction et division - le sujet peut inverser la stratégie de lecture : si un élève a mieux mémorisé la table de multiplication que celle de division, et doit vérifier 42 :7 = 6, il peut lire l'égalité de droite à gauche et vérifier que 6x7 = 42.

Une tâche de vérification conduit, secondairement, à des effets non observables dans une tâche de production. Ont notamment été identifiés : (1) l'effet de positivité selon lequel une égalité correcte (e.g., 7x8 = 56) sera traitée plus rapidement qu'une égalité non trop grossièrement incorrecte (e.g., 7x8 = 63); (2) l'effet d'interférence, ou de confusion, inter-opérations selon lequel une égalité incorrecte (e.g., 7+8 = 53) sera traitée plus rapidement

qu'une égalité, incorrecte aussi, mais pour laquelle la réponse proposée correspond au résultat d'une autre opération arithmétique appliquée aux deux termes de l'opération (e.g., 7+8 = 56). Ce dernier effet est particulièrement intéressant pour éprouver des modèles théoriques de représentation en mémoire. Il a ainsi donné lieu, consécutivement à sa première mise en évidence par Winkelman et Schmidt (1974), à une pléthore de recherches : Frenck-Mestre et Vaid (1993), Lemaire et al. (1991, 1994), pour n'en citer que quelques-unes récentes. Toutefois, l'intensité des effets obtenus dans ces différentes recherches ne semble pas suffisante pour leur attribuer une grande pertinence pédagogique.

La tâche de vérification ne peut donc pas être simplement décrite comme une tâche de production à laquelle s'ajouterait un stade de comparaison avec la réponse proposée (Campbell, 1987; Zbrodoff & Logan, 1990), même s'il existe quelques évidences contraires (Timmers & Claeys, 1990) discutées dans Fischer (1992, p.103).

1.1.3 *La validité de la tâche de vérification*

Les contraintes matérielles de la tâche de production nous ont conduit à choisir une tâche de vérification dans la méthode Juste-Faux exposée ci-après. Les spécifités de la tâche de vérification, que nous venons de résumer, nous ont alors incité, lors de nos premières observations suivant la méthode Juste-Faux, à étudier la corrélation avec la forme d'évaluation habituellement en vigueur dans les classes concernées. Malheureusement, nous n'avons qu'assez rarement trouvé des notes de calcul mental, a fortiori d'évaluation de la connaissance des tables élémentaires. Nous avons donc souvent dû nous résigner à calculer une corrélation avec des performances en calcul écrit, voire en mathématique plus généralement. Le bilan de ces analyses de corrélations - nous en rapportons une précise dans l'illustration de la partie 2 suivante - est qu'elles sont significatives et bonnes, sans être parfaites. Un tel résultat nous paraît tout à fait rassurant pour la méthode : d'une part, du fait que la corrélation est significativement bonne, il montre que l'on ne mesure pas n'importe quoi; d'autre part, du fait que la corrélation n'est pas parfaite, il montre aussi que l'évaluation ne fait pas double-emploi avec d'autres formes d'évaluation.

Par ailleurs, même si la tâche de vérification ne mesure peut-être pas exactement la même chose que la tâche de production, et peut conduire à la mise en jeu de traitements différents, on peut remarquer qu'elle ne devrait pas entraver la réussite d'un élève qui ne saurait que produire les faits numériques élémentaires. En effet, un tel élève pourra toujours utiliser cette connaissance pour répondre correctement, même si, à défaut de stratégie plus efficiente, sa vitesse de réponse pourrait laisser à désirer. En outre, la tâche de vérification possède un intérêt propre, et croissant avec l'utilisation de plus en plus répandue des calculettes et ordinateurs.

1.2 La méthode Juste-Faux (Fischer, 1988a)

1.2.1 *Phase de production des données*

La méthode repose principalement sur la mesure des TR. De manière plus précise, elle s'appuie sur un programme destiné aux élèves en fin d'école élémentaire (CE à CM), en début de collège ou en Section d'Etudes Spécialisées. La figure 1 présente le décours temporel d'une question.

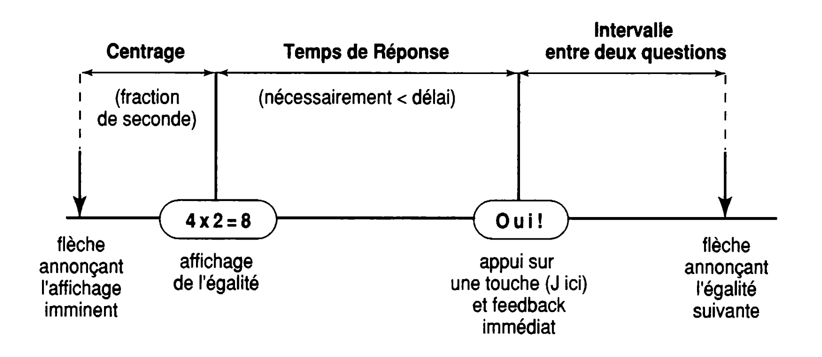


FIGURE 1

Décours temporel d'une question.

Outre le choix méthodologique principal que constitue la mesure des TR, et celui d'une tâche de vérification que nous venons de commenter, d'autres choix, dont l'importance est loin d'être négligeable, ont été faits subséquemment. Nous les explicitons rapidement.

1. La limitation du temps de réponse : la figure 1 fait apparaître que les élèves n'ont qu'un **délai limité** pour répondre. Ce délai peut être choisi dans le programme d'initialisation entre 2 et 9 secondes.
2. La nature - additions, soustractions, multiplications, ou divisions - des opérations à proposer. Vu l'objectif d'évaluation et de comparaison (inter-opérations), les **quatre** types d'**opérations** sont également représentés.
3. Les **deux modalités** de présentation : lors de l'une des deux passations exigées par la méthode, l'élève se voit proposer 14 additions, puis 14 soustractions, puis 14 multiplications et enfin 14 divisions (modalité **REG**roupement), en étant prévenu, avant chaque bloc de 14, de la nature de l'opération suivante; au cours de l'autre passation, il se voit proposer les opérations «mélangées» (modalité **N**on-**REG**roupement), en ignorant la nature de l'opération suivante.
4. Les égalités : le programme d'initialisation de Juste-Faux permet de choisir entre **plusieurs jeux d'égalités**. Ces égalités ne «sortent» jamais des tables élémentaires. La figure 2 montre qu'elles sont partagées en deux niveaux : sans définir précisément ces derniers, on peut

dire que le niveau 1 concerne les faits arithmétiques portant sur des petits nombres (e.g., 2x3 = 6), alors que le niveau 2 concerne des faits portant sur des nombres plus grands (e.g., 6x8 = 48).

5. Les deux passations : chaque session d'évaluation comprend deux passations. On en «profite» pour varier la modalité de présentation (REG ou NREG) et pour présenter un autre sous-jeu d'égalités.

1.2.2 *Phase de traitement des données*

Une fois que l'ensemble des élèves d'une classe ou d'un groupe a effectué les deux passations, des programmes de traitement permettent d'analyser leurs TR signés, i.e. pour lesquels un code précise s'ils correspondent à une réponse correcte, incorrecte ou non-réponse dans le délai imparti. Ces programmes permettent, entre autres, de ranger les élèves ou les égalités d'après le nombre de réussites ou d'après le TR. Mais le traitement le plus immédiatement lisible et, surtout, qui tente d'intégrer l'ensemble des mesures et des variables, est l'image de la classe ou du groupe.

Pour construire une image (cf., la figure 2 qui en présente trois), nous utilisons la technique de visualisation de l'ensemble Juste-Faux (Fischer, 1988a). Une case élémentaire correspond à une opération de niveau donné. Si un élève, repérable par son numéro, a échoué à plus de 2 vérifications sur les 14 concernant cette opération et ce niveau, sa case correspondante reste blanche; sinon, on calcule son TRu (Temps de Réussite, i.e. un TR limité à des réponses correctes) pour les 12 meilleures performances et, suivant le résultat, on lui attribue un niveau de gris (voir légende de la figure 2). Comme le noir reflète les meilleures performances, les lignes les plus noircies sont celles correspondant aux élèves les plus performants; en conséquence, le progrès de la classe ou du groupe se traduit par une image qui se noircit ou, à tout le moins, devient plus foncée, au fil des sessions.

2. ILLUSTRATION DE LA MÉTHODE JUSTE-FAUX

2.1 Conditions de l'observation

L'observation rapportée dans cet article utilise une préversion du logiciel Juste-Faux (Fischer, 1988a). Elle concerne une classe de CM2 (5ème année d'école en France) de 25 élèves qui n'a subi aucune modification d'effectif en cours d'année. L'expérience est intégrée au travail scolaire habituel et se fait, bien entendu, en accord et avec la coopération de la maîtresse.

Le plan d'évaluation consiste en trois mesures recouvrant toute l'année scolaire. Une mesure se fait au cours d'une session, chaque session comportant deux passations (voir ci-avant) espacées d'environ une semaine. La session 1 s'est déroulée en septembre, une ou deux semaines après la rentrée; la session 2, fin janvier, et la session 3, en juin, deux ou trois semaines avant les grandes vacances. Le délai de réponse est fixé à 5 secondes.

Les élèves ont été informés, en septembre, de l'ensemble du projet et, lors de chaque session, et aussi d'une passation à l'autre, l'expérimentateur insistait sur le fait qu'il s'agit certes de répondre vite, mais qu'il faut répondre correctement avant tout. Les passations se faisaient en général par groupes de 6 élèves dans la salle informatique équipée d'un NanoRéseau (Léanord) standard en présence du seul expérimentateur.

2.2 Résultats

Comme annoncé dans la partie 1.1.3, nous avons procédé, pour cette classe, à un calcul de la corrélation entre les performances à Juste-Faux et à une évaluation scolaire de la maîtrise des opérations arithmétiques des élèves au premier semestre. Cette corrélation ne porte que sur 24 élèves car un des élèves était absent lors de l'évaluation (écrite et collective) en classe. Elle est bonne et significative sans être parfaite : pour la session 1, le coefficient de corrélation r est de.49 [$t(22) = 2.63$, $p = .015$]; pour la session 2, $r = .58$ [$t(22) = 3.33$; $p = .003$]. La part de variance des résultats scolaires expliquée par la performance à Juste-Faux lors des sessions 1 et 2 est, respectivement, de 24% et de 34%.

La technique de visualisation, décrite de manière générale ci-avant, a conduit aux images comparatives de la figure 2. Sur ces images, chaque élève repérable par son numéro de 01 à 25, se voit attribuer une ligne en fonction de sa performance en septembre, cette ligne étant alors conservée pour les deux autres sessions.

2.3 Discussion

La visualisation de la figure 2 confirme que les élèves sont loin de connaître uniformément et définitivement les faits numériques élémentaires en début de CM2 (5ème et dernière année d'école élémentaire), ni même à la fin. Pour les différences entre élèves, les images sont d'une telle netteté qu'elles n'ont pas besoin d'être commentées. Pour les différences entre opérations, on peut par exemple voir que les soustractions de niveau 1 sont beaucoup moins bien maîtrisées que les multiplications de même niveau.

Par ailleurs, la visualisation donne une impression de progrès, général et important dans le sens où il concerne tous les élèves et toutes les opérations et niveaux. Elle suggère clairement qu'aucun élève performant en début d'année (haut de l'image de septembre), n'avait atteint un «plafond» l'empêchant de progresser par la suite. En outre, on peut remarquer que les 8 élèves du haut de l'image de septembre - les seuls à avoir au moins une case gris foncé et moins de 3 cases blanches -, se détachent visuellement dès le début de l'année scolaire et encore plus nettement à la fin. On peut voir aussi que certains élèves - les numéros 20 et 15 notamment - ont quand même plus progressé que d'autres - les numéros 24 et 14 par exemple. Enfin, si on ne se laisse pas trop impressionner par le noir dominant de la session de juin, on peut entrevoir un ralentissement du progrès dans la deuxième moitié de l'année, peut-être en partie dû à une démobilisation de certains élèves à la veille des grandes vacances.

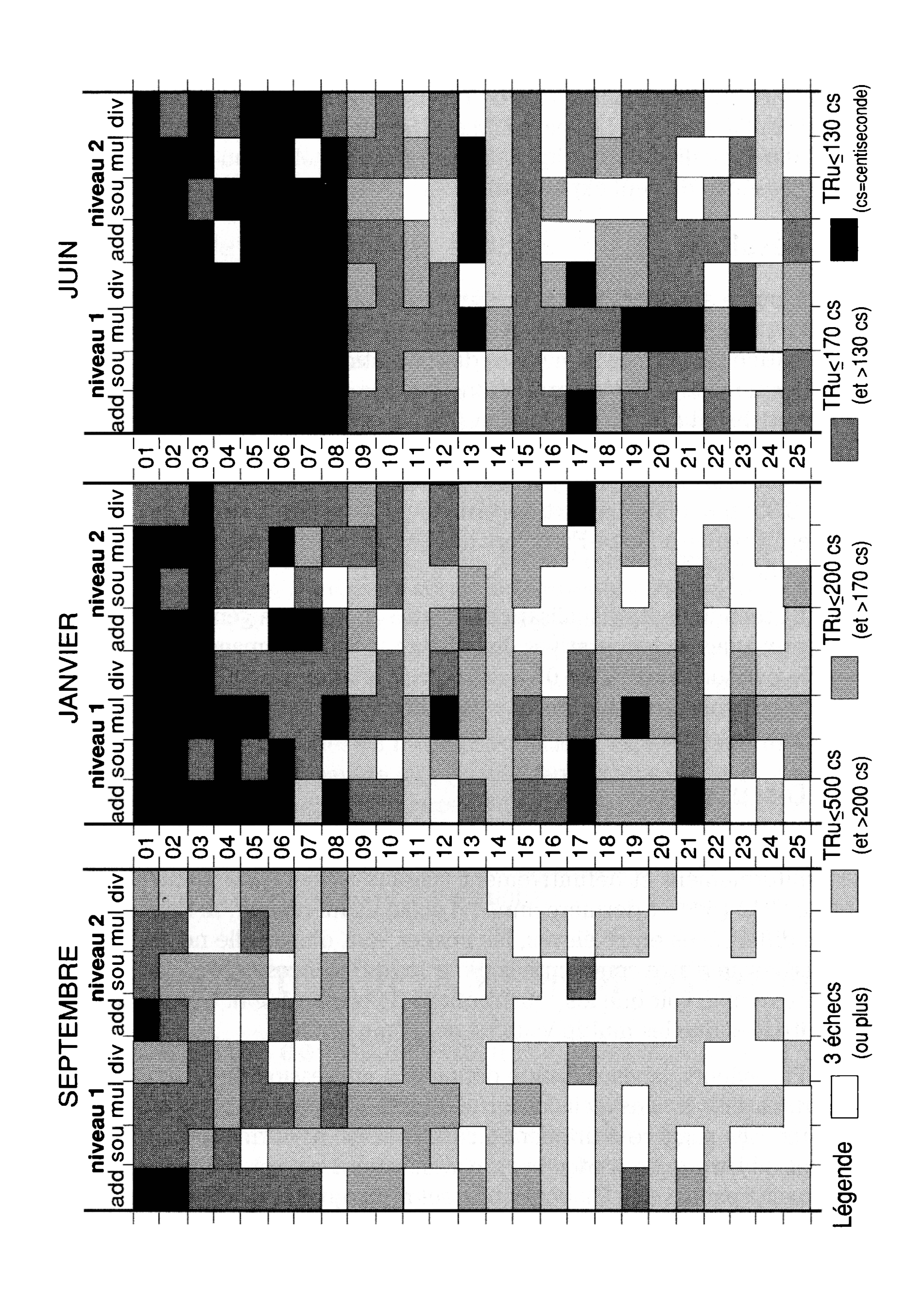


FIGURE 2

Résultats des élèves lors des trois évaluations.

Quelle est l'origine de ce progrès massif ? On peut, là aussi, penser à un effet «grandes vacances» : la première session s'étant déroulée peu après les grandes vacances, il est probable que certains élèves avaient oublié en partie des faits numériques qu'ils connaissaient par coeur avant les vacances. Mais, outre cet artefact, l'ampleur du progrès nécessite d'autres explications. Les deux principales semblent la maturation et la pratique. Nous ne commenterons pas la première, mais nous nous attarderons un peu sur la seconde en essayant de préciser la manière dont pourrait s'exercer son influence.

D'abord, la pratique pourrait conduire à des changements dans les stratégies de calcul ou de vérification. Pour la stratégie de vérification, les élèves peuvent, s'ils ont mieux compris l'inversion des opérations, partir du résultat proposé dans les opérations inverses et vérifier, par exemple, 9+3 = 11 à la place de 11-3 = 9. La présente observation, contrairement à celle de Fischer (1988b) pour la division, n'étaye guère une telle évolution : les opérations inverses (divisions et soustractions), qui auraient dû le plus profiter d'un tel changement de stratégie, n'ont apparemment pas plus progressé que les opérations directes (additions et multiplications). Une explication possible est que, dans cette classe, les élèves avaient déjà suffisamment compris l'inversion opératoire en début d'année pour que l'on n'assiste plus à une évolution importante en cours d'année.

Pour la stratégie de calcul, les élèves peuvent soit changer de procédure, soit passer d'une méthode reconstructive s'appuyant sur une mémoire procédurale à une méthode reproductive s'appuyant sur une mémoire déclarative (cf. Fischer, 1992, pour la distinction procédural/déclaratif). Un changement de procédure consiste, par exemple, à passer d'un comptage de un en un (e.g., pour 9+5, dire 10, 11, 12, 13, **14**, en parcourant parallèlement les doigts d'une main) à un passage par 10 plus économique (e.g., pour 9+5, dire 9+1, **10**, plus 4, **14**). Un changement de méthode consiste à passer d'une méthode procédurale, comme celles que nous venons de décrire, à une méthode déclarative, i.e. à une récupération en mémoire à long terme déclarative d'une unité - phrase, proposition ou image - notée {+, 9, 5, 14} dans Fischer (1992). Par exemple, dans la présente observation, l'égalité 9x3 = 27 est celle qui a le plus gagné en vitesse en cours d'année. On peut donc penser qu'elle a été reconstruite, par additions successives ou en parcourant le début de la table de 7, en début d'année (juste après les vacances !) par certains élèves, alors qu'elle était connue par coeur par quasiment tous les élèves en fin d'année scolaire.

2.4 Conclusions

D'abord, nous aimerions suggérer un élément de réponse à une question que soulève cette observation : Comment des élèves, qui semblent connaître par coeur des faits numériques rudimentaires comme 4+2 = 6, 2+3 = 5, 4x2 = 8, ... dès la fin de la première ou deuxième année d'école, peuvent-ils encore progresser aussi nettement (et plus que ce que semblent expliquer de simples facteurs maturationnels) ? Nous suggérons que l'évaluation scolaire

«classique» n'est pas assez précise et rigoureuse pour repérer tous les élèves qui ne connaissent pas les faits élémentaires par coeur. En particulier, le délai de réponse n'est pas contrôlé de manière rigoureuse, et le TR n'est quasiment jamais mesuré. Une illustration de ce constat est constituée, dans la présente observation, par l'élève n° 15 sur la visualisation de la figure 2, première de la classe y compris en mathématiques, qui connaissait très peu de faits par coeur et n'avait pas été repérée par la maîtresse : dans toutes les évaluations scolaires, elle avait probablement le temps et la capacité de reconstruire le résultat !

Ensuite, nous conclurons cette partie en soulignant, par référence à Ehrlich et Florin (1989), que la méthode Juste-Faux semble remplir deux fonctions :

1. elle permet aux élèves de «voir» leurs progrès et, en conséquence, de les inciter à en faire, à l'image des grands champions qui «ne pensent pas qu'ils vont devenir les meilleurs tout de suite; ils s'entraînent beaucoup, tous les jours. Ce qui les intéresse c'est de voir qu'ils progressent, qu'ils font mieux chaque semaine que la semaine précédente » (p.45).
2. elle permet aux enseignants un diagnostic rapide, précis et peu coûteux, évitant ainsi «que le maître passe tout un temps (quelques mois) à bien comprendre comment fonctionne chaque enfant; puis à repérer ultérieurement tous les changements qui interviennent dans ce fonctionnement » (ibid., p.47).

3. APPLICATIONS DE LA MÉTHODE DES TR

3.1 A l'élaboration d'un modèle du développement numérique

Une autre préversion du logiciel Juste-Faux, que nous avons utilisée sur 210 élèves en fin d'école élémentaire (dernier trimestre du CM2), a conduit à deux observations intéressantes.

1. Les quatres opérations sont loin d'être uniformément connues : ce sont surtout les opérations additives qui conduisent à des performances (réussites et TR) médiocres (Fischer, 1987). Par exemple, une «égalité» comme 11-3= 9 n'a été réussie que par 54,3% des élèves (à peine plus que le hasard), avec un TRu ≈ 2,71 s; ou encore, 9+7 = 15 a conduit à 59,5% de réussite, avec un TRu≈ 2,69 s. Si l'on ajoute qu'une égalité comme 5x2 = 10 a conduit à 93,3% de réussite, avec un TRu deux fois moindre, on voit bien qu'il existe une différence de difficulté considérable entre des égalités toutes qualifiées d'élémentaires. En outre, on peut noter que les élèves présentant un retard scolaire (quelle qu'en soit l'origine) se sont révélés, électivement, en difficulté sur une égalité comme 9+7 = 15 : ils ont non seulement été

plus lents, mais surtout leur pourcentage de réussite n'a été que de 39% (contre 68% pour les autres élèves).

2. Les soustractions s'opposent aux multiplications sur un axe procédural / déclaratif dégagé par l'Analyse Factorielle des Correspondances (AFC) de Fischer et Pluvinage (1988; voir figure 3). Il est probable que les premières conduisent très souvent à une reconstruction du résultat, grâce à une procédure stockée en mémoire procédurale, alors que les secondes conduisent davantage à une récupération du résultat en mémoire déclarative. Autrement dit, les élèves **savent** seulement **comment** on fait pour trouver 11-3, alors que, pour 5x2 (ou, plus typiquement, pour 7x7) ils **savent que** c'est 10 (ou 49).

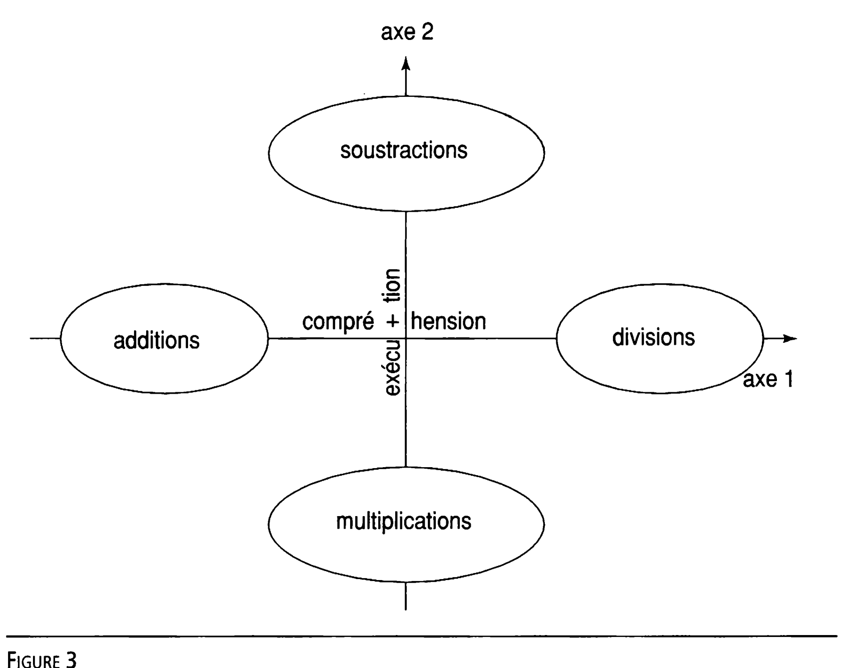


FIGURE 3

Situation des quatres opérations sur les axes compréhension et exécution.

En considérant alors les soustractions comme le prototype des connaissances procédurales, et les multiplications comme celui des connaissances déclaratives, nous avons pu émettre plusieurs hypothèses, non triviales, qui se sont confirmées (Fischer, 1992). Les connaissances procédurales sont plus sensibles à l'amorçage (indirect) que les connaissances déclaratives. En revanche, ce sont les connaissances déclaratives qui sont le plus sensibles à l'effet de positivité : la supériorité des égalités positives ou justes (e.g., 7x7 = 49) sur les égalités fausses (e.g., 7x7 = 42) est plus accusée pour la multiplication que pour la soustraction.

Mais cette distinction entre connaissances procédurales et déclaratives, pour fondamentale qu'elle soit, n'est peut être pas l'aspect développementalement le plus intéressant du modèle PDup (Procédural-Déclaratif-**u**nitisation-**p**océduralisation) que nous avons élaboré. En effet, comme le suggère le sigle PDup, nous postulons aussi l'existence de deux mécanismes d'échange et d'interaction entre ces deux types de connaissance. L'**unitisation** permet de former de nouvelles unités en mémoire déclarative, directement ou à partir de la mise en oeuvre de connaissances procédurales. Par exemple, l'élève qui trouve 7+3 grâce à la procédure de surcomptage (8, 9, **10**), peut ensuite former une unité, notée {+,7,3,10}, qui associe les données et le résultat. La **procéduralisation**, quant à elle, permet l'inclusion de connaissances déclaratives dans des procédures. En prolongeant l'exemple précédent, le même élève pourra alors inclure cette unité dans une procédure de «passage de la dizaine», comme 7+5 = 10+2 = 12. Notons d'ailleurs que cette dernière procédure inclut non seulement l'unité {+,7,3,10}, mais aussi, subséquemment, l'unité {+,3,2,5}.

Enfin, un autre aspect pédagogiquement intéressant du modèle, est qu'il superpose deux formes d'apprentissage aux deux formes de connaissances ou de mémoires. L'apprentissage **déclaratif** est direct et abrupt : nous pouvons encoder presque instantanément une nouvelle unité. Néanmoins, une nouvelle unité, pour subsister à long terme, devra être **consolidée**. L'apprentissage **procédural** est, en revanche, plutôt lent et graduel. Il est le fruit de l'exercice : on apprend en faisant pour reprendre la formule d'Anderson (1983). Avec l'exercice, les procédures peuvent s'**automatiser**. Nous avons ainsi une distinction, théoriquement fondée, entre deux mécanismes souvent présents dans le discours pédagogique mais rarement bien différenciés : la consolidation et l'automatisation.

3.2 A l'étude des différences individuelles

Widaman, Little, Geary et Cormier (1992) ont étudié transversalement le développement de l'addition mentale. Leur étude 1 a porté sur 123 élèves de 2ème, 4ème et 6ème années ayant répondu à 140 additions dans une tâche de vérification; en outre, 8 scores à des subtests du SAT (Stanford Achievement Test) sont aussi disponibles pour ces élèves. Leur étude 2 a porté sur 163 étudiants de 2ème, 4ème, 6ème année et du collège (19.8 ans en moyenne) soumis à des additions mentales simples (80) et complexes (80); en outre, ils ont aussi été soumis à des tests de facilité numérique et de vitesse de perception.

Les auteurs arrivent à plusieurs conclusions :

– il y a des différences individuelles substantielles pour l'âge où se produit la transition du comptage vers la récupération : alors qu'une large proportion d'élèves de 2ème année a déjà recours à la récupération pour répondre aux additions, une importante minorité d'élèves de

6ème année, et même d'étudiants de collège, a toujours recours à des processus de comptage;

- le degré d'automatisation des différents processus qui sous-tendent l'addition mentale (encodage des chiffres, recalcul ou récupération de la somme correcte, retenue à reporter dans la colonne suivante) n'est pas le même pour tous;
- si l'on prend pour indice la vitesse de récupération pour les sujets récupérant les faits et la vitesse d'incrémentation pour ceux qui les comptent, les différences individuelles dans l'addition sont fortement reliées aux mesures de niveau et de capacité.

Si ces conclusions, notamment la dernière, confirment donc l'intérêt de la mesure des TR et lui confèrent en outre une validité externe, en revanche les enseignants ne trouveront rien de révolutionnaire dans les «implications éducatives» que Widaman et Little (1992, pp.247-8) essaient de tirer de cette recherche (entre autres). En effet, Widaman et Little soulignent : (1) que l'influence de l'éducation sur les additions élémentaires se fait surtout sentir dans les premières années; (2) qu'il faudrait ajuster l'enseignement de chaque problème aux capacités de l'enfant individuel; (3) que le niveau scolaire est un meilleur indice que l'âge chronologique pour mesurer le développement des additions. Tout ceci est presque évident, ou pour le (2), contenu, au moins implicitement, dans les grandes théories de l'apprentissage, en tout cas dans celles que nous avons exposées dans un document récent (Fischer, 1993b).

3.3 À l'étude de groupes particuliers d'élèves

3.3.1 *L'étude des sujets MD*

Geary, Brown et Samaranayake (1991) ont entrepris une étude longitudinale sur 26 sujets normaux et 12 sujets *Mathematically Disabled* (MD) de 1ère et 2ème année d'école. Au moment de la 1ère mesure, les élèves ont eu à calculer 40 additions simples extraites des 56 que l'on obtient en enlevant les 0, 1 et doubles. Dix mois après, la tâche d'addition a été administrée une seconde fois. La tâche des élèves est ici une tâche de production.

Chez les élèves normaux, Geary et al. observent un recours plus important à la récupération et moins important au comptage, aussi bien qu'une amélioration des vitesses de comptage et de récupération. En revanche, chez les élèves MD aucun changement fiable dans le mixage des stratégies ou dans la vitesse d'exécution du comptage et des récupérations en mémoire n'a été observé. De manière précise, d'une session à l'autre, l'utilisation du comptage a baissé de 23 % et celle de la récupération a augmenté de 12% chez les Normaux, alors que chez les MD on observe une certaine stabilité, voire un pattern évolutif inverse : le comptage a augmenté de 6%, alors que la récupération a baissé de 1%.

A partir de l'observation de Geary et al. (1991), et aussi par référence à Howell, Sidorenko et Jurica (1987) qui ont montré (pas trop scientifiquement à notre avis) qu'un apprentissage sévère (*drill*) et extensif de faits de base n'améliore pas la récupération, Geary (1993) suggère l'existence de deux déficits fonctionnels distincts. Ces deux déficits semblent suivre des trajectoires développementales différentes. Le premier déficit, l'utilisation de procédures arithmétiques développementalement immatures et une grande fréquence d'erreurs procédurales, semble médiatisé par un retard développemental dans l'acquisition de connaissances conceptuelles qui sous-tendent l'utilisation des procédures. Le second déficit implique une difficulté plus fondamentale, i.e. qui ne disparaît pas avec le développement, dans la représentation et la récupération de faits arithmétiques de la mémoire à long terme sémantique (ou déclarative comme nous préférons dire). En se basant sur des études comparatives de jumeaux homozygotes et dizygotes, Geary suggère même que ce deuxième déficit pourrait être davantage héréditaire que le premier.

3.3.2 *Les sujets HyperActifs*

Zentall et Smith (1993), après Zentall (1990) qui avait étudié des sujets LD (*Learning Disabled*) et avec déficits attentionnels (ADD : *Attention Deficit Disorder*), ont étudié 57 garçons sans désordre, 22 HyperActifs (HA) non agressifs et 13 HA agressifs. Les opérations arithmétiques étudiées, sur des nombres à 1 chiffre seulement, sont : l'addition, la soustraction et la multiplication, avec la particularité que la soustraction peut être posée dans l'ensemble Z des relatifs (e.g., «2 - 7 = »).

La tâche est une tâche de production : les élèves doivent taper les réponses au clavier, aussi vite que possible, et appuyer sur Entrée à la fin, cette dernière frappe arrêtant le chrono. Les auteurs ont essayé de contrôler le facteur vitesse de frappe en introduisant préalablement une tâche de recopie d'assemblages du type «- 8 =» qui a une fonction à la fois d'entraînement et de mesure de la seule vitesse de frappe.

Bien que cette recherche soulève certains problèmes méthodologiques - notamment la pratique curieuse de la soustraction avec des élèves recrutés de la 2ème à la 5ème année d'école - on ne peut manquer de noter que les auteurs soulignent fortement l'intérêt et la pertinence de la mesure des vitesses de l'addition. Par exemple, dès l'introduction, ils soulignent :

- que Zentall (1990) a montré que la vitesse de calcul des additions et des soustractions est inférieure à celle des normaux chez les sujets LD et chez les sujets ADD ;
- la valeur heuristique des facteurs sous-tendant la vitesse de réponse en maths : la vitesse des calculs simples s'est en effet avérée le meilleur prédicteur de la reconnaissance en lecture et de la résolution de problèmes.

Puis, en conclusion de leurs observations et de celles antérieures avec des sujets souffrant de désordres plus sévères (Zentall, 1990), ils remarquent que :

- la vitesse plus lente de l'addition pourrait être un marqueur d'une infériorité des performances en addition, et des performances scolaires plus généralement, chez les jeunes souffrant d'une hyperactivité accompagnée d'agressivité;
- la vitesse de réponse aux faits d'addition pourrait être une meilleure mesure que beaucoup de tâches de laboratoire ou de tests normatifs.

Enfin, dans l'abstract, ils écrivent que «les conclusions dérivées à partir de ces découvertes, et de travaux antérieurs, indiquent que la vitesse de l'addition pourrait être un marqueur de dysfonctionnement scolaire et social».

Conclusion

Dans cet article, nous avons envisagé un problème méthodologique majeur posé par la mesure des TR en situation scolaire (partie 1.1). Nous avons expliqué comment la méthode Juste-Faux tente de le surmonter (partie 1.2). Cette dernière a conduit à de nombreuses observations semblables à celle rapportée dans la partie 2. L'observation précise rapportée nous a alors permis de montrer l'utilisation pédagogique de la méthode, notamment à des fins d'évaluation diagnostique. Enfin, à propos de la méthode des TR plus généralement, nous avons rapporté les grands traits du modèle de développement numérique qu'elle nous a permis de mettre au point (partie 3.1), ainsi que des applications récentes à l'étude des différences individuelles (partie 3.2) et à celle de groupes (MD ou HA) particuliers (partie 3.3).

Dans l'ensemble, ces applications récentes ont confirmé - parfois nettement (voir Zentall et al. ci-avant) l'intérêt diagnostique, à des fins de caractérisation ou de différenciation, de la mesure des TR à des faits numériques élémentaires.

Nous conclurons (évidemment !) de manière au moins aussi positive quant à l'application de la méthode Juste-Faux. Nous ne reviendrons pas sur les intérêts de la méthode, que nous avons dégagés dans les parties 2.3 et 2.4. Nous soulignerons cependant encore, à propos du modèle PDup qu'elle nous a permis d'élaborer (cf. la partie 3.1), que ce modèle prend une nouvelle dimension au moment où certains élèves de Piaget (e.g., Mounoud, 1994) postulent que c'est l'existence simultanée de deux systèmes de connaissances en parallèle qui constituerait la dynamique du développement. En effet, et en accord avec Mounoud, nous débouchons alors sur la conclusion générale que le moteur du développement, selon la célèbre formule de Piaget, ne serait pas l'action comme il l'avait soutenu mais les rapports dialectiques qui s'instaurent entre deux systèmes de connaissances.

Mais tous ces intérêts de la méthode Juste-Faux, et des autres mesures analogues des TR aux faits numériques élémentaires, ne manquent pas de soulever une question intrigante à laquelle nous tenterons de répondre en cette fin d'article : pourquoi une telle évaluation ne s'est-elle pas (encore ?) imposée dans les classes ?

Nous passerons sur les problèmes pratiques bien qu'ils jouent vraisemblablement un rôle majeur. Nous insisterons davantage sur deux autres raisons.

La première est d'ordre institutionnel. Dans les programmes officiels français (cf., MEN, 1991), on demande une mémorisation de la table d'addition en 1ère année d'école (Cours Préparatoire) et de celle de multiplication en 2ème et 3ème année (Cours Elémentaire). Mais on se garde bien de préciser les critères d'une telle mémorisation. En effet, si l'on voulait préciser ces critères, on serait vraisemblablement obligé d'introduire, d'une manière ou d'une autre, les TR.

La seconde, qui est en partie à la fois une cause et une conséquence de la première, est que la distinction entre une connaissance déclarative vs procédurale, quels que soient les qualificatifs utilisés (e.g., récupération vs reconstruction), est loin de s'être imposée, pour les tables élémentaires, à tous les esprits. On pense encore trop souvent, et le rappel des programmes ci-avant l'illustre bien, que les tables sont mémorisées précocement, uniformément et définitivement par les élèves. Toutes nos observations expérimentales et analyses montrent qu'il n'en est rien (voir, par exemple, Fischer, 1987; voir aussi la partie 2.3). D'où vient alors cette fausse idée ? Nous l'expliquons par le fait que dans les recherches psychologiques on n'interroge souvent que des sujets étudiants, volontaires, répondant plusieurs milliers de fois à des additions élémentaires, ... ; on arrive alors à un résultat - les sujets récupèrent les faits arithmétiques dans un réseau ou une table - dont la validité écologique est douteuse. En outre, et en complément, les deux modèles basés sur l'observation de jeunes élèves les plus influents ont, trop lourdement nous semble-t-il, insisté sur la récupération en mémoire, au détriment d'autres stratégies (e.g., le calcul réfléchi : cf. Brissiaud, Clerc & Ouzoulias, 1992) : Ashcraft (1982) en suggérant que c'est vers la fin de la troisième année d'école que les enfants passent d'une méthode algorithmique (e.g., le comptage) à une récupération en mémoire déclarative; Siegler & Shrager (1984) en donnant la priorité à la récupération en mémoire dans leur modèle initial de choix de stratégie. D'ailleurs, pour étayer ce dernier point de vue, nous terminerons en citant une recherche récente de Geary, Frensch et Wiley (1993). Ces derniers semblent considérer comme une «découverte» - en tout cas ils la mettent en valeur dans leur conclusion - l'observation que tous les étudiants universitaires gradés ne sont pas versés en arithmétique. De manière précise, leur étude sur la soustraction, jointe à celle semblable sur l'addition de Geary et Wiley (1991), suggèrent que moins de la moitié des étudiants semble

maîtriser l'arithmétique de base si on la définit comme une récupération automatique en mémoire à long terme des faits de base et des faits «en colonne» [1] (e.g., 32-9).

1 Récupérer des faits «en colonne» en mémoire à long terme c'est appliquer la technique opératoire de l'opération posée en recourant aux faits de base.



LA LECTURE



Évaluer les difficultés de lecture chez l'adulte

Christine Jourdain, Daniel Zagar et Bernard Lété

Introduction

Proposer une évaluation-diagnostic des apprentissages, c'est proposer une évaluation qui permette de déterminer l'origine d'une (ou de plusieurs) difficulté(s) observée(s) dans un domaine considéré. Ainsi, en lecture, l'évaluation-diagnostic doit être capable non pas de donner simplement une estimation du niveau d'habileté d'un lecteur en terme de vitesse de lecture et/ou de compréhension (ce que font généralement les tests de lecture), mais d'identifier l'origine d'une éventuelle inefficience, en indiquant le plus précisément possible le niveau de performance d'une personne pour les différents processus qui permettent à la lecture de bien se dérouler (Jourdain, 1989; Seymour, 1986; Zagar, 1992b). En 1990, la région Bourgogne, dans le cadre de la lutte contre l'illettrisme et d'un contrat de plan avec l'Etat, nous accordait un financement pour mettre au point une évaluation de ce type pour le lecteur adulte. En construisant le logiciel d'évaluation-diagnostic des Capacités Cognitives du Lecteur Adulte (Zagar, Jourdain & Lété, 1993; Zagar, Jourdain & Lété, à paraître), notre objectif était donc de créer un logiciel qui donne des indications sur les capacités cognitives qui sous-tendent l'activité du lecteur et sur leur niveau de fonctionnement.

Pour mener à bien un tel projet, il est nécessaire d'une part de bénéficier d'un cadre théorique de référence qui permette de faire le point sur les différentes capacités impliquées en lecture. Ce cadre nous est fourni par les travaux en psycholinguistique et fera l'objet de la première partie de cet exposé. D'autre part, il faut proposer des épreuves adaptées à l'évaluation du fonctionnement des différents processus impliqués en lecture et préalablement définis. Là encore, les travaux en psycholinguistique permettent de résoudre un certain nombre de problèmes, ce qui fera l'objet d'une seconde partie. Enfin, dans une dernière partie nous présenterons les résultats de l'étalonnage auquel nous avons procédé auprès de 160 adultes, en montrant plus particulièrement quelles sont les différences observables entre des lecteurs plus ou moins efficients.

1. CADRE THÉORIQUE

Depuis les années soixante, la lecture est considérée comme un ensemble d'étapes de traitement de l'information. Si les modélisations varient d'un auteur à l'autre (pour une revue cf. Zagar, 1992a), il existe néanmoins un certain nombre de points communs : au départ, l'information est extraite de la page; à la fin, cette information est intégrée à l'ensemble de nos connaissances et permet l'élaboration d'une représentation mentale du texte. La figure 1 présente une illustration des étapes qui permettent cette élaboration. Globalement, on peut distinguer trois étapes nécessaires à la compréhension d'un texte : la reconnaissance des mots (niveaux 1 et 2 sur le modèle), l'analyse syntaxique (niveau 3) et l'analyse sémantique (niveau 4).

1.1 La reconnaissance des mots

Reconnaître un mot c'est mettre en correspondance un signal visuel (le mot écrit) et une représentation interne de ce signal. Les psycholinguistes font l'hypothèse que ces représentations internes sont stockées au sein d'une sorte de dictionnaire : le lexique mental, qui se construit progressivement au cours de l'apprentissage de la langue. Cependant cet appariement entre signal visuel et représentation interne ne s'effectue pas de façon directe.

En premier lieu (niveau 1 sur la Figure 1) interviennent les processus d'extraction des informations contenues dans la page, parmi lesquels il faut ranger les traitements perceptifs, notamment la reconnaissance des lettres et l'analyse de la chaîne de caractères (i.e. le codage des lettres en fonction de leur position; Ellis & Young, 1990). A ce niveau, les lettres vont activer des codes internes : les codes graphémiques (Bigsby, 1988; Coltheart, 1981; Evett & Humphreys, 1981).

Une fois ces codes graphémiques activés, il existerait deux façons d'activer une entrée lexicale (i.e. la représentation interne d'un mot; Coltheart, 1978).

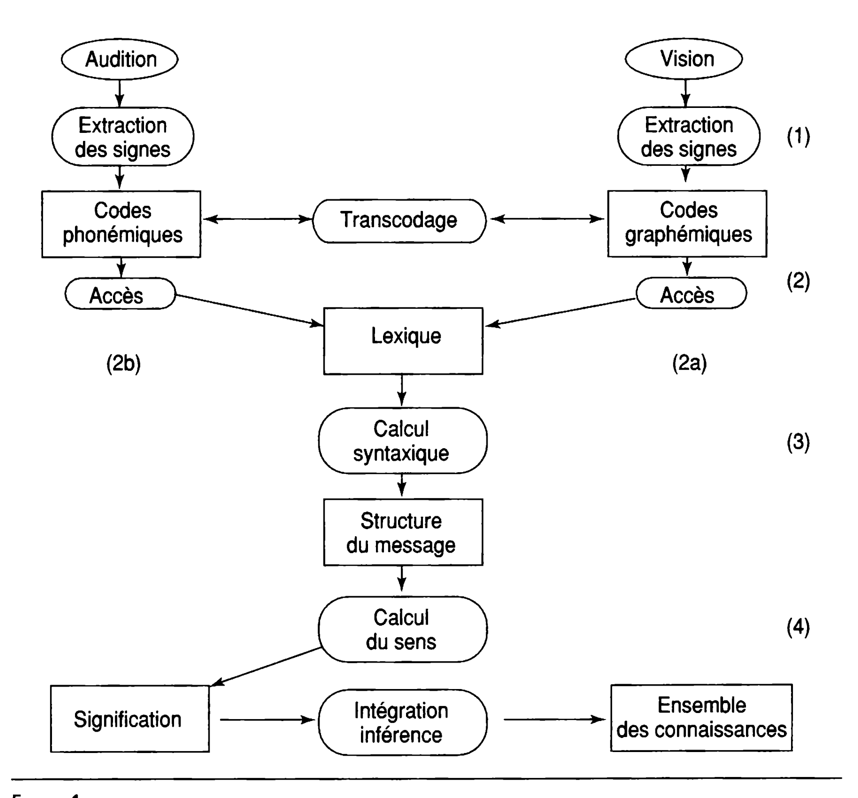


FIGURE 1

Modélisation des étapes de traitement de l'information impliquées pendant la lecture

La première (2a sur la figure 1) consiste en l'appariement direct des codes graphémiques avec les codes stockés au sein du lexique mental : on parle alors de voie directe ou d'adressage (Forster, 1976; McClelland & Rumelhart, 1981).

La seconde (2b sur la figure 1) consiste, à partir des codes graphémiques, en l'activation de codes phonémiques (par l'intermédiaire d'une procédure mettant en oeuvre des règles de correspondance entre graphème et phonème) et en l'assemblage de ces codes phonémiques. Dans ce cas, la reconnaissance du mot s'effectue par appariement de cette représentation phonologique assemblée avec une représentation phonologique contenue dans le lexique mental (Segui, 1991). On parle ici de voie indirecte ou d'assemblage. Cette voie joue sans doute un rôle important lors de l'apprentissage de la lecture, en permettant à l'enfant d'associer des représentations graphiques, nouvelles pour lui, à des représentations auditives familières (Barron, 1986; McCusker, Hillinger & Bias, 1981; Meyer, Schvaneveldt & Ruddy, 1974; Stanovich, 1982).

1.2 L'analyse syntaxique (niveau 3)

Pour comprendre un texte, le lecteur doit non seulement identifier les mots mais aussi déterminer quelles sont les relations qu'ils entretiennent entre eux (Mitchell, 1987). Autrement dit, il s'agit d'articuler les items lexicaux entre eux, de repérer leur rôle fonctionnel : sujet, verbe,..., de les associer en unités de plus haut niveau : syntagme, proposition,...(Zagar, 1992a).

1.3 L'analyse sémantique (niveau 4)

Une fois la structure de la phrase établie, le lecteur procède au calcul de la signification, i.e. à l'intégration des informations au reste du texte. Il va également intégrer ces informations à l'ensemble de ses connaissances et cela en fonction de ses connaissances préalables.

Si la finalité de toute lecture est de comprendre, l'activité de lecture, comme l'indique la Figure 1, ne peut se résumer à la compréhension. Par conséquent, on peut logiquement envisager que chacun des processus impliqués dans la lecture puisse être responsable ou à l'origine d'une difficulté en lecture. Cependant, du point de vue du fonctionnement, ces processus ne sont pas équivalents puisqu'ils nécessitent plus ou moins de ressources attentionnelles. En effet, compte tenu, d'une part, de la rapidité avec laquelle le lecteur expert effectue l'ensemble des opérations qui mènent à la compréhension (la vitesse moyenne de lecture varie entre 240 et 400 mots/minute selon le type de texte; Rayner & Pollatsek, 1987), compte tenu, d'autre part, du fait que les ressources attentionnelles dont dispose le système de traitement sont limitées (Kahneman, 1973), il faut envisager qu'un certain nombre d'opérations sont automatisées (Laberge & Samuels, 1974).

Le paradigme de Stroop (1935) a permis de mettre en évidence le caractère automatique (i.e. irrépressible) de la reconnaissance des mots. Lorsque l'on présente à une personne une planche sur laquelle sont inscrits des noms de couleurs (rouge, bleu, vert,...) dans des encres de couleur différente (par exemple, rouge écrit en bleu...) et qu'on lui demande de nommer la couleur de l'encre des mots, on observe (par rapport à la tâche de dénomination des mots) un ralentissement au niveau de la vitesse de prononciation, des hésitations et parfois des erreurs. Tout se passe comme si la personne ne pouvait s'empêcher de lire le mot, comme si la réponse au mot s'imposait face à la réponse à la couleur.

Chez le lecteur expert, l'ensemble des opérations menant à la reconnaissance des mots serait donc automatisé, permettant ainsi d'allouer les ressources attentionnelles aux processus de haut niveau (étapes chargées du calcul du sens). A l'opposé, le lecteur débutant mobiliserait ses ressources attentionnelles vers les premières étapes du traitement, ce qui entraverait le processus de compréhension. De fait, on a pu montrer que l'habileté à décoder les mots explique en grande partie le niveau de compétence en lecture (Schvaneveldt, Ackerman & Semlear, 1977; Stanovich, 1982, 1986). Par ailleurs, dans la majorité des cas, il semble que les difficultés de lecture proviennent d'un déficit au niveau des processus menant à la reconnaissance des mots (cf. les travaux sur les dyslexies développementale et acquise), notamment par un défaut d'automatisation (ou une détérioration dans le cas des dyslexies acquises) des processus menant à cette reconnaissance (Bruck, 1990; Stanovich, 1982; Vellutino, 1979). Par ailleurs, ces processus sont spécifiques à la lecture contrairement aux processus de compréhension qui sont impliqués dans d'autres tâches cognitives (compréhension du langage parlé). C'est pourquoi ECCLA propose essentiellement une évaluation-diagnostic des processus qui conduisent à la reconnaissance des mots. Ces processus ont été regroupés en étape de traitement visuel (i.e. activation des codes graphémiques et analyse de la chaîne de caractères), étape de traitement phonologique (i.e. activation des codes phonémiques, assemblage et activation des entrées lexicales par les codes phonologiques), et étape lexicale (i.e. activation directe des entrées lexicales et activation des concepts au sein du lexique).

2. TECHNIQUES, MATÉRIEL ET ÉTALONNAGE

Une fois définies les différentes étapes impliquées dans la compréhension d'un texte, se pose le problème des techniques qui permettent d'évaluer le fonctionnement de chacune de ces étapes de traitement. Si l'on y réfléchit, on s'aperçoit que l'acte de lecture ne donne lieu à presqu'aucun comportement observable (en dehors de celui de l'activité oculaire et de la lecture à voix haute). Les psycholinguistes ont mis au point des techniques expérimentales susceptibles de révéler l'activité de processus mentaux parfaitement inobservables en situation naturelle (pour une revue cf., par exemple, Rayner & Pollatsek, 1989). La plupart de ces techniques utilise le temps nécessaire à

l'accomplissement d'une tâche simple comme indicateur du fonctionnement d'une étape de traitement (Sternberg, 1969). Le principe est que les étapes de traitement sont consommatrices de temps et que tout dysfonctionnement devrait se traduire par un ralentissement des processus engagés. Autrement dit, plus le traitement est automatisé, plus la réponse est rapide.

2.1 Techniques

On peut distinguer deux types de mesures chronométriques utilisées en recherche fondamentale : les techniques dites à temps de réaction et celles dites à temps libre (Zagar, 1992a).

Les techniques à temps de réaction consistent à demander au sujet d'effectuer le plus rapidement et le plus précisément possible une tâche simple. Par exemple, dans le cas d'une décision lexicale, la personne devra décider si la suite de lettres qui apparaît à l'écran est ou n'est pas un mot de la langue française. Le temps de réaction (TR) correspond à l'intervalle de temps entre l'apparition à l'écran de la suite de lettres et le moment où le sujet donne sa réponse (en appuyant sur une touche déterminée du clavier). Ce temps de réaction est supposé refléter la vitesse à laquelle la reconnaissance du mot est effectuée et donc les processus cognitifs qui conduisent à cette reconnaissance. Outre le fait qu'en soi un TR est déjà un indicateur du traitement cognitif effectué par le sujet, on peut également manipuler différentes caractéristiques du matériel (longueur de la chaîne de caractères, fréquence des mots,...) et donc comparer la façon dont s'effectuent les traitements.

Les techniques à temps libre consistent à mesurer de façon continue le temps que met un lecteur pour lire chaque segment (mot, syntagme,...) d'un texte. Par exemple, le paradigme APS II (Auto-Présentation Segmentée version 2; Lété, Zagar & Pynte, 1994) permet au lecteur de commander l'apparition de chaque mot du texte à l'écran. Pour faire apparaître le mot, le lecteur utilise une touche du clavier : l'appui de cette touche provoque l'apparition du mot à l'écran, le lâcher sa disparition, il faut appuyer de nouveau pour faire apparaître le mot suivant. Pour chaque mot, nous disposons donc d'un temps de lecture (en millisecondes) décomposable en un temps d'appui (temps pendant lequel le mot reste affiché à l'écran et qui reflète, selon les auteurs, au moins, les processus d'extraction de l'information) et un temps de lâcher (temps pendant lequel l'écran reste vide) et qui reflète les processus subséquents à l'accès au lexique.

Les épreuves utilisées pour l'évaluation-diagnostic des processus menant à la reconnaissance des mots vont donc s'appuyer sur le recueil de telles mesures chronométriques.

2.2 Matériel[1]

Il est constitué de trois ensembles d'épreuves :

1. dix épreuves à temps de réaction visant à évaluer le fonctionnement des étapes (visuelle, phonologique -indirecte- et lexicale -directe) menant à la reconnaissance des mots, auxquelles nous avons ajouté;
2. deux épreuves sur un matériel non linguistique, comprenant deux tâches à temps de réaction qui déterminent la vitesse de réaction du sujet;
3. deux épreuves de lecture en APS II (technique à temps libre).

2.2.1 *Les épreuves testant l'analyse visuelle*

Nous avons sélectionné quatre épreuves pour estimer l'activation des codes graphémiques et l'analyse d'une chaîne de caractères.

Codage abstrait des lettres (Posner & Mitchell, 1967). Dans cette épreuve, le sujet doit décider si les deux lettres présentées simultanément à l'écran sont identiques indépendamment de la façon dont elles sont typographiées (deux majuscules -A A-, deux minuscules -a a- ou une majuscule et une minuscule -A a). Les lettres ne sont pas réversibles (p/q), la lettre écrite en majuscule a une forme différente de la lettre écrite en minuscule (A/a et non C/c). Pour les réponses négatives (c X) les deux lettres ont été choisies de façon à ce qu'elles soient rarement confondues (moins de 1% d'erreur selon la matrice de confusion de Paap, Newsome, McDonald & Schvaneveldt, 1982). Après 12 items d'entraînement, le sujet doit effectuer cette décision pour 36 items expérimentaux présentés chacun deux fois et répartis en 6 blocs.

Jugement d'identité de lettres (Seymour, 1986). On demande au sujet de déterminer si une suite de lettres est constituée uniquement par la même lettre (par exemple : CCCYCCC, réponse négative, vs CCCCCCC, réponse positive). Les suites sont de 3, 7 ou 11 lettres majuscules. Dans le cas où une autre lettre est présente, celle-ci peut être soit très différente (CY ou MB) soit très proche (KX ou PR) selon la matrice de confusion de Paap et al. (1982). Cette épreuve comporte 24 items d'entraînement répartis en 2 séries et 72 items expérimentaux répartis en 6 séries.

Jugement de similitude de 2 ensembles non-prononçables de lettres (Seymour, 1986). On présente simultanément l'une à côté de l'autre deux suites de 5 consonnes en majuscule. Le sujet doit déterminer si les deux suites sont identiques. Les suites diffèrent par le nombre de lettres modifiées (une, BRTQP BLTQP, vs toutes, BRTQP CSLMN), par le type de modification apportée (substitution d'une lettre, BRTQP BLTQP, ou inversion de deux lettres, BRTQP BTRQP) ainsi que par la position de la modification apportée (de la première à la cinquième lettre). Dans le cas d'une modification portant sur une seule lettre, cette lettre est remplacée par une lettre donnant lieu à peu

1 Annexe 1 pour une synthèse des épreuves et des exemples.

de confusion (Paap & al., 1982). Le sujet dispose de 22 items d'entraînement répartis en 2 blocs et de 88 items expérimentaux répartis en 8 blocs.

Jugement de similitude de 2 ensembles prononçables de lettres (Seymour, 1986). On présente simultanément l'une à côté de l'autre deux suites de 5 lettres en majuscule. Le sujet doit déterminer si les deux suites sont identiques. Les suites diffèrent par le nombre de lettres substituées (une, CABOL CATOL, vs toutes, CABOL DUFIS) et par la position de la lettre remplacée (de la première à la cinquième lettre). Le sujet dispose de 16 items d'entraînement répartis en 2 blocs et de 48 items expérimentaux répartis en 6 blocs.

2.2.2 *Les épreuves testant l'analyse phonologique*

Trois épreuves permettent de tester l'activation des codes phonémiques et des entrées lexicales par le biais de ces codes.

Décision lexicale phonologique. Le sujet doit décider si un pseudo-mot est homophone d'un mot de la langue française (par exemple : phoule - foule- vs phabi). Vingt-quatre pseudo-mots, de 2 à 3 syllabes, ont été construits à partir de mots dont la fréquence d'occurrence dans la langue est élevée (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 3.7; Content, Mousty & Radeau, 1990) et de mots dont la fréquence d'occurrence dans la langue est basse (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 1.7; Content & al., 1990). La transformation d'un mot en pseudo-mot s'effectue en modifiant soit un graphonème fréquent (i, f,...), soit un graphonème rare (y, ph,...) de la première syllabe (Veronis, 1986). Le sujet dispose de 16 items d'entraînement répartis en 2 blocs et de 48 items expérimentaux répartis en 6 blocs.

Décision phonologique à choix forcé 1. Dans cette épreuve, on présente simultanément deux non-mots l'un en dessous de l'autre. Le sujet doit décider si les deux items se prononcent de la même façon. Les items sont de une syllabe (en moyenne 4 lettres et 3 phonèmes) ou deux syllabes (les items de 2 syllabes sont construits à partir des items d'une syllabe, en moyenne 7.5 lettres et 5 phonèmes) et peuvent être homographes et homophones (naubir naubir) ou simplement homophones (naubir neaubir). Le sujet dispose de 16 items d'entraînement répartis en 2 blocs et de 64 items expérimentaux répartis en 8 blocs.

Décision phonologique à choix forcé 2. Dans cette épreuve, le sujet doit décider si un mot et un non-mot présentés simultanément l'un en dessous de l'autre se prononcent de la même façon (monsieur mecieu). Les mots peuvent être de fréquence d'occurrence élevée (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 3.7; Content & al., 1990) ou basse (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 1.6; Content & al., 1990), régulier ou irrégulier (Content, 1991). Les non-mots sont construits en régularisant les mots. Le sujet dispose de 12 items d'entraînement et de 96 items expérimentaux répartis en 8 blocs.

2.2.3 *Les épreuves testant l'étape lexicale*

Quatre épreuves de décision lexicale (jugement d'appartenance à la langue française) permettent de tester la vitesse d'activation des entrées lexicales; une cinquième épreuve teste l'activation des concepts.

Amorçage orthographique. On demande au sujet de décider si un mot (cible) précédé par une amorce appartient à la langue française. La cible est fréquente (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 3.7; Content & al., 1990) ou rare (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 1.7; Content & al., 1990). L'amorce est constituée par le mot écrit en lettres majuscules, le temps d'affichage de l'amorce (SOA, Stimulus Onset Asynchrony) peut être court (50 ms) ou long (100 ms). Le sujet dispose de 24 items d'entraînement répartis en 2 blocs et de 72 items expérimentaux répartis en 6 blocs.

Décision lexicale 1. Dans cette épreuve les mots sont plus ou moins fréquents (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 4 pour les mots fréquents et de 1.8 pour les mots rares; Content & al., 1990). En ce qui concerne les non-mots, ils peuvent être soit non prononçables (constitués exclusivement de consonnes), soit prononçables (constitués de consonnes et voyelles). Les mots comme les non-mots sont de différentes longueurs (de 3 à 9 lettres). Le sujet dispose de 24 items d'entraînement répartis en 2 blocs et de 84 items expérimentaux répartis en 7 blocs.

Décision lexicale 2. Dans cette épreuve les mots (identiques à ceux utilisés dans la décision phonologique à choix forcé 2) sont réguliers ou irréguliers (Content, 1991) et de fréquences différentes (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 3.8 pour les mots fréquents et de 1.6 pour les mots rares; Content & al., 1990). En ce qui concerne les non-mots, ils peuvent être soit non prononçables (constitués exclusivement de consonnes), soit prononçables (constitués de consonnes et voyelles). Le sujet dispose de 12 items d'entraînement et de 96 items expérimentaux répartis en 8 blocs.

Décision lexicale 3. Dans cette épreuve les mots sont plus ou moins fréquents (le logarithme en base 10 de la fréquence pour 100 millions est en moyenne de 5.3 pour les mots fréquents et de 3.5 pour les mots rares; Content & al., 1990) et de différentes classes (classe ouverte : verbes, substantifs, adjectifs; classe fermée : prépositions, conjonctions, adjectifs possessifs,...; Segui, Mehler, Frauenfelder & Morton, 1982). En ce qui concerne les non-mots, ils peuvent être soit non prononçables (constitués exclusivement de consonnes), soit prononçables (constitués de consonnes et voyelles). Le sujet dispose de 20 items d'entraînement répartis en 2 blocs et de 80 items expérimentaux répartis en 8 blocs.

Catégorisation sémantique (Collins & Loftus, 1975; Cordier, 1980). Dans cette épreuve le sujet doit décider si un item écrit en minuscule appartient à la catégorie affichée en majuscule sur l'écran (OISEAU moineau). Les

items peuvent être soit typiques (moineau pour la catégorie OISEAU) soit atypiques (autruche), soit proches (papillon) soit éloignés (chaise) de la catégorie. Le sujet dispose de 8 items d'entraînement et de 64 items expérimentaux répartis en 8 blocs.

2.2.4 *Les épreuves non linguistiques*

Il s'agissait, pour les deux épreuves concernées, de recueillir un temps de réaction moteur sur un matériel qui ne soit pas linguistique.

Temps de réaction simple. Dans cette épreuve, le sujet doit appuyer sur une des touches du clavier lors de l'apparition d'un signe en forme de coeur. L'intervalle de temps précédent l'apparition du coeur à l'écran varie de 250 à 2500 ms. Cette épreuve comporte 20 items d'entraînement répartis en 2 blocs et 30 items expérimentaux répartis en 3 blocs.

Temps de réaction à choix forcé. Le sujet doit décider si le signe qui apparaît à l'écran est un coeur ou un carré. Cette épreuve comporte 10 items d'entraînement et 30 items expérimentaux répartis en 3 blocs. Le temps de réaction moyen pour les réponses « oui » correctes constitue le temps de réaction de base du sujet. Ce temps sera systématiquement retiré du temps de réaction moyen des épreuves linguistiques.

2.2.5 *Les épreuves de lecture*

Deux épreuves de lecture, l'une de textes, l'autre de mots isolés, se déroulent au moyen de la technique d'Auto-Présentation Segmentée II (Lété, Zagar & Pynte, 1994).

Lecture de textes. Les sujets lisent en mot à mot 18 textes de 5 propositions et de 27 à 43 mots (Lété et al., 1994). Ils disposent de 3 textes d'apprentissage et effectuent un jugement de titre à la fin de chaque texte.

Lecture de mots. Après 3 listes d'apprentissage, les sujets lisent en mot à mot 19 listes constituées de 5 à 13 mots. Les listes sont construites de façon à intégrer des mots de fréquence et de longueur différentes, ainsi que des mots lus en contexte (lecture de textes). A la fin de chaque liste, les sujets effectuent un jugement de synonymie sur le dernier mot de la liste.

2.3 Etalonnage

2.3.1 *Sujets*

Cent-soixante sujets ont passé l'ensemble des épreuves. Ils ont été répartis en trois groupes selon leur niveau scolaire. Le groupe 1 comprend 57 personnes de niveau second cycle du secondaire et plus (baccalauréat et plus), âgées de 18 à 28 ans (moyenne = 21.5 ans), 34 sujets féminins et 23 masculins; le groupe 2 est composé de 65 personnes de niveau fin de premier cycle du secondaire (CAP, BEP, BEPC), âgées de 18 à 47 ans (moyenne = 23 ans), 10 sujets féminins et 55 masculins; enfin le groupe 3 est formé de 38 per-

sonnes de niveau fin d'études primaires, âgées de 16 à 57 ans (moyenne = 27 ans), 13 sujets féminins et 25 masculins, qui participaient pour la majorité d'entre eux à des stages de remise à niveau en français et mathématique. Ces sujets n'ont été soumis à aucun test de lecture préalable.

2.3.2 *Procédure*

Pour les épreuves à temps de réaction, les sujets effectuent leur décision en utilisant deux touches du clavier. La main dominante est associée aux réponses «oui».

La passation des épreuves se déroule dans un ordre prédéterminé pour tous les sujets en quatre séances d'environ une heure, à raison d'au maximum une passation par demi-journée. La première session comprenait les épreuves suivantes (entre parenthèses est précisée l'abréviation du nom de l'épreuve) : temps de réaction simple (TRS), temps de réaction à choix forcé (TRC), codage abstrait des lettres (COD), lecture de textes (LMM); la seconde session comprenait de nouveau l'épreuve de temps de réaction à choix forcé (TRC), suivi du jugement d'identité de lettres (JIL), du jugement de similitude de deux ensembles prononçables de lettres (SM2), de la décision lexicale 1 (DL1), de la décision phonologique à choix forcé 1 (PH1) et de la lecture de mots (LHC). Lors de la troisième session, les épreuves étaient : TRC, décision lexicale 2 (DL2), amorçage orthographique (AOR), catégorisation sémantique (SEM) et décision phonologique à choix forcé 2 (PH2). Enfin dans la dernière session, les sujets effectuaient les épreuves suivantes : TRC, décision lexicale 3 (DL3), jugement de similitude de deux ensembles non-prononçables de lettres (SM1) et décision lexicale phonologique (PHO).

3. RÉSULTATS

En présentant quelques résultats, nous voudrions essayer de répondre à deux questions :

1. quelles différences peut-on mettre en évidence pour les trois groupes quant à la vitesse de résolution des tâches et donc au niveau de l'activation des processus qu'elles testent ? Pour répondre à cette question, nous nous appuierons sur les résultats moyens de chaque groupe.
2. quelles différences peut-on observer entre lecteurs d'un même groupe ? Nous analyserons pour ce faire les résultats des sujets en fonction des classes effectuées lors de l'étalonnage (i.e. en fonction de leur position sur la distribution des notes centrées réduites).

3.1 Vitesse de lecture et compréhension

Les trois groupes se différencient-ils du point de vue des épreuves de lecture ?

Comme le montre le Tableau 1, la lecture est plus rapide dans le cas d'une lecture de textes (LMM) que dans celui de listes de mots (LHC), respectivement 622 millisecondes par mot (ms/mot) vs 788 ms/mot (F (1, 157) = 124.03, p <.001). Cette vitesse diminue en fonction du niveau de qualification : plus les lecteurs sont qualifiés et moins ils passent de temps sur chaque mot (434 ms/mot pour le groupe 1; 695 ms/mot pour le groupe 2 et 1128 ms/mot pour le groupe 3; F (2, 157) = 52.22, p <.001).

TABLEAU 1

Vitesse moyenne de lecture, pourcentage du temps d'appui et pourcentage de bonnes réponses pour chacun des trois groupes

	Groupes			
	1	**2**	**3**	**Moy**
LMM vitesse	387 (130)	626 (291)	967 (437)	622 (364)
%TA	50 (11)	54 (14)	54 (17)	53 (14)
%BR	96 (5)	92 (9)	88 (12)	83 (9)
LHC vitesse	**481 (162)**	**765 (339)**	**1290 (618)**	**788 (489)**
%TA	**47 (13)**	**60 (17)**	**60 (19)**	**56 (17)**
%BR	**87 (6)**	**83 (7)**	**78 (12)**	**92 (9)**

Note : écarts-types entre parenthèses.

Vitesse en millisecondes par mot, %TA = pourcentage de temps de lecture par mot; %BR = pourcentage de bonnes réponses, LMM = lecture de textes; LHC = lecture de listes de mots.

Par ailleurs, le ralentissement de la vitesse de lecture est d'autant plus important en l'absence de contexte que le niveau scolaire est faible (+94 ms/mot pour le groupe 1, +139 ms/mot pour le groupe 2 et +323 ms/mot pour le groupe 3; F (2, 157) = 15.34, p <.001). L'effet facilitateur du contexte est donc plus important pour les lecteurs les moins efficients (pour une revue, cf. Becker, 1985) ce qui apparaît conforme aux résultats de West et Stanovich (1982).

Au niveau de la compréhension, le jugement de titre est moins bien réussi que le jugement de synonymie (83 vs 92 % de bonnes réponses; F (1, 157) = 135.21, p <.001).

Les sujets du groupe 1 réussissent globalement mieux (respectivement pour les groupes 1, 2 et 3 : 91, 87 et 83 % de bonnes réponses; F (2, 157) = 16.88, p <.001). En revanche, contrairement à la vitesse de lecture, le nom-

bre de bonnes réponses (par ailleurs très élevé) ne varie pas de façon statistiquement significative d'une épreuve à l'autre en fonction du niveau scolaire ($F < 1$).

En ce qui concerne la façon dont s'effectue la lecture, on peut noter que le temps d'affichage des mots à l'écran (%TA) est proportionnellement plus important lors de la lecture de listes de mots que lors de la lecture de textes (le temps d'appui représente respectivement 56 et 53% du temps de lecture d'un mot; $F (1, 157) = 15.45$, $p <.001$). La part du temps d'appui varie en fonction du niveau scolaire : les lecteurs de faible niveau gardent le mot plus longtemps à l'écran (en moyenne : 48 % pour le groupe 1; 57 % pour les groupes 2 et 3; $F (2, 157) = 7.32$, $p <.001$). En l'absence de contexte, l'augmentation du temps d'appui est d'autant plus importante que les sujets sont des lecteurs lents (-3 % pour le groupe 1, +6 % pour les groupe 2 et 3; $F (2, 157) = 13.31$, $p <.001$). Ce résultat apparaît conforme aux résultats de Olson, Kliegl, Davidson et Foltz (1985) qui ont distingué deux types de lecteurs : des lecteurs «laborieux» et des lecteurs «explorateurs», les premiers prenant plus de temps pour assurer la reconnaissance des mots, les seconds se limitant à encoder suffisamment d'informations visuelles afin de poursuivre ultérieurement les traitements subséquents.

Afin d'évaluer plus précisément l'influence du contexte, nous avons calculé l'effet de facilitation associé à la présentation des mots en contexte, en analysant les temps d'appui et de lâcher pour 105 mots communs aux deux épreuves de lecture (cf. Tableau 2).

TABLEAU 2

Effet (en ms) de facilitation du contexte sur les temps d'appui (TA) et de lâcher (TL) pour chacun des trois groupes.

	Groupes			
	1	**2**	**3**	**Moy**
TA	-43 (73)	-128 (123)	-230 (219)	-122 (155)
TL	-66 (86)	-26 (111)	-72 (203)	-51 (132)
Moyenne	-54	-77	-151	

Note : écarts-types entre parenthèses

La facilitation associée à la présentation des mots en contexte apparaît en moyenne plus importante pour les temps d'appui que pour les temps de lâcher (respectivement -122 et -51 ms; $F (1, 157) = 24.60$, $p <.001$), pour les lecteurs les plus lents que pour les lecteurs rapides (-151 ms pour le groupe 3, -77 ms pour le groupe 2 et -54 ms pour le groupe 1; $F (2, 157) = 12.56$, $p <.001$). Cependant, cette facilitation ne se manifeste pas de la même façon

pour les trois groupes. Pour le groupe 1, elle a tendance à être plus importante sur les temps de lâcher que sur les temps d'appui ($z = 1.53$, $p < .06$), alors que pour les groupes 2 et 3 elle intervient massivement sur les temps d'appui ($F (2, 157) = 10.95$, $p < .001$; groupe 2, $z = 4.96$, $p < .01$; groupe 3, $z = 3.26$, $p < .01$). Les lecteurs les moins efficients éprouvent donc une difficulté au niveau de l'extraction des informations écrites lorsqu'ils ne peuvent s'appuyer sur le contexte; ce qui conforte le fait que le bon lecteur n'a pas besoin de contexte pour identifier efficacement les mots (West et Stanovich, 1978, 1982).

Dans un premier temps, le niveau scolaire peut donc être considéré comme un bon estimateur du niveau de lecture.

Les sujets du groupe 1 (niveau baccalauréat et plus) constituent un groupe de lecteurs rapides (387 et 481 ms/mot en LMM et LHC) et «bons compreneurs» (96 et 87 % de bonnes réponses pour les jugements de titre et de synonymie), qui équilibrent leurs ressources cognitives entre l'extraction et l'intégration de l'information (respectivement 50 et 47 % d'appui en LMM et LHC).

Par comparaison, les sujets du groupe 3 (niveau fin d'études primaires) forment un groupe de lecteurs très lents (967 et 1290 ms/mot, soit trois fois plus que les lecteurs du groupes 1), plus faibles au niveau de la compréhension bien que les performances restent élevées (88 et 78 % de bonnes réponses) et qui ont besoin de plus de temps pour extraire les informations du mot affiché à l'écran (54 et 60 % d'appui en LMM et LHC) et donc pour reconnaître ce mot.

Enfin, les sujets du groupe 2 (niveau fin du premier cycle du secondaire) peuvent être considérés comme constituant un groupe intermédiaire au niveau de la compréhension (respectivement 92 et 83 % de bonnes réponses en LMM et LHC). Au niveau de la vitesse de lecture (respectivement, 626 et 765 ms/mot en LMM et LHC), ils apparaissent plus proches des lecteurs du groupe 1 (+239 et +284 ms en LMM et LHC) que de ceux du groupe 3 (respectivement +341 et +525 ms). En revanche, en ce qui concerne les processus d'extraction de l'information, ils sont comparables à ceux du groupe 3 (54 et 60 % d'appui en LMM et LHC).

3.2 Épreuves non-linguistiques

L'analyse des résultats aux épreuves non-linguistiques : temps de réaction simple (TRS) et temps de réaction à choix forcé (TRC) permet de mettre en évidence des différences entre les trois groupes au niveau des temps de réaction (cf. Tableau 3).

TABLEAU 3

Temps de réaction moyen (en ms) pour les épreuves de temps de réaction simple (TRS) et de temps de réaction à choix forcé (TRC) pour chacun des trois groupes

	Groupes			
	1	2	3	Moy
TRS	289 (32)	339 (86)	392 (128)	334 (93)
TRC	378 (39	432 (70)	473 (89)	423 (76)

Le temps de réaction moyen est moins élevé en TRS qu'en TRC (340 vs 428 ms; F (1, 157) = 231.83, p <.001). En revanche, si globalement le temps de réaction augmente quand le niveau scolaire diminue (F (2, 157) = 24.87, p <.001), la différence entre les deux épreuves est identique pour les trois groupes (respectivement 89, 93 et 81 ms; F < 1). Autrement dit, dans le cas d'épreuves non linguistiques, le surcroît de difficulté associé à la décision (s'agit-il oui ou non de...., ici s'agit-il oui ou non d'un coeur ?) est le même quel que soit le niveau d'étude.

3.3 Processus impliqués dans la reconnaissance des mots

Afin d'obtenir un indicateur «au plus près» du temps nécessaire à la mise en oeuvre des processus impliqués dans les étapes visuelle, phonologique ou lexicale, nous avons soustrait, pour chaque sujet, son temps de réaction moyen dans l'épreuve TRC (temps de réaction à choix forcé) à son temps de réaction moyen pour chacune des épreuves testant une étape particulière. Cette mesure (ici appelé coût) correspond donc au temps nécessaire à l'exécution d'une tâche moins un temps de décision et de programmation motrice (temps qui ne nous intéresse pas ici). Ce sont les résultats pour les coûts associés aux processus que nous présentons ici et cela en fonction des différentes étapes préalablement définies.

3.3.1 *Etape visuelle*

Existe-t-il des différences entre les groupes dès les épreuves testant les processus d'activation des codes graphémiques et d'analyse d'une chaîne de caractères ?

Le Tableau 4 présente pour chaque groupe le coût associé aux différentes épreuves d'évaluation de l'étape visuelle .

Nous pouvons observer dans un premier temps une progression dans la difficulté associée aux tâches : COD (codage abstrait des lettres) est réussi plus rapidement que JIL (jugement d'identité de lettres; z = 17.12, p <.01) et

JIL plus rapidement que SM2 (jugement de similitude de 2 ensembles de lettres prononçables; z = 12.08, p <.01). Par ailleurs, en ce qui concerne les épreuves de jugement de similitude de deux ensembles de lettres prononçables (SM1) ou non-prononçables (SM2), on peut noter que, si le coût augmente en fonction du groupe (F (2, 152) = 30.54; p <.001) et en fonction de l'épreuve (F (1, 152) = 643.76; p <.001), en revanche, la facilitation apportée par la prononçabilité des suites de lettres n'apparaît pas significativement différente pour les trois groupes, le coût de la tâche étant divisé par deux (Groupe 1, -629 ms; Groupe 2, - 656 ms; Groupe 3, -752 ms : F (2, 152) = 2.13; p <.12).

Dans toutes les épreuves, le coût associé à la tâche (et donc aux processus d'activation des codes graphémiques et/ou d'analyse de la chaîne de caractères) diminue en fonction du niveau scolaire : les sujets du groupe 1 étant plus rapides que ceux des groupes 2 et 3 et ceux du groupe 2 plus rapides que ceux du groupe 3 (cf. Annexe 2). Donc, à l'exception de l'épreuve de codage abstrait des lettres pour laquelle les groupes 2 et 3 ne se différencient pas (COD; z =.83, p <.21), il existe, comme en lecture, une progression entre les trois groupes : plus le niveau scolaire est faible, plus le coût associé aux épreuves est élevé. Cependant, contrairement aux résultats obtenus en lecture, le groupe 2 semble plus proche du groupe 3 que du groupe 1 : excepté pour SM1, les différences entre les coûts sont plus importantes entre les groupes 1 et 2 (pour COD : +35 ms; JIL : +102 ms; SM1 : +254 ms; SM2 : +227 ms) qu'entre les groupes 2 et 3 (respectivement; +15, +71, +256, +160 ms).

TABLEAU 4

Coût moyen (en ms) pour les épreuves testant l'étape visuelle pour chacun des trois groupes

	Groupes			
	1	**2**	**3**	**Moy**
COD	133 (40)	168 (87)	183 (90)	159 (77)
JIL	298 (69)	400 (128)	471 (183)	380 (144)
SM1	1140 (305)	1394 (390)	1650 (557)	1368 (466)
SM2	511 (125)	738 (289)	898 (338)	696 (298)

Note : écarts-type entre parenthèses.
Codage abstrait des lettres (COD), Jugement d'Identité de Lettres (JIL), Jugement de Similitude de 2 ensembles de lettres non-prononçables (SM1) ou prononçables (SM2).

Ainsi, dès la mise en oeuvre des processus d'activation des codes graphémiques et d'analyse d'une chaîne de caractères, les groupes 2 et 3 obtiennent des coûts supérieurs à ceux du groupe 1. Il semble donc qu'une partie

des difficultés de lecture peut être associée à des difficultés observables dès le traitement des lettres.

3.3.2 *Étape phonologique*

Qu'en est-il des différences au niveau des processus mis en oeuvre dans les tâches testant l'étape phonologique ?

Pour les trois épreuves concernées, on peut observer (cf. Tableau 5), comme précédemment, un effet du groupe (F (2, 138) = 45.92, p <.001). Ainsi, dans toutes les épreuves le groupe 1 réussit plus rapidement que le groupe 2, qui lui même réussit plus rapidement que le groupe 3 (cf. Annexe 2), à l'exception toutefois de la décision lexicale phonologique (PHO) pour laquelle les groupes 2 et 3 ne se différencient pas (z = 1.41, p <.08).

TABLEAU 5

Coût moyen (en ms) pour les épreuves testant l'étape phonologique pour chacun des trois groupes.

	Groupes			
	1	**2**	**3**	**Moy**
PHO	394 (101)	593 (222)	672 (264)	547 (264)
PH1	531 (152)	761 (259)	991 (370)	730 (311)
PH2	584 (153)	866 (300)	1030 (422)	795 (336)

Note : écarts-types entre parenthèses.
Décision lexicale phonologique (PHO), décision phonologique à choix forcé 1 (PH1) et 2 (PH2)

On observe également un effet de la tâche (F (2, 276) = 106.91, p <.001) ainsi qu'une interaction entre le groupe et la tâche (F (4, 276) = 3.87, p <.004) : PHO donne lieu à des coûts moins élevés que la décision phonologique à choix forcé 1 qui concerne 2 non-mots (PH1; pour le groupe 1, z = 5.77, p <.01; pour le groupe 2, z = 3.88, p<.01; pour le groupe 3, z = 4.07, p <.01), qui elle-même donne lieu à des coûts moins élevés que la décision phonologique à choix forcé 2 qui concerne un mot (régulier ou irrégulier) et un non-mot (PH2; pour le groupe 1, z = 1.85, p <.05; pour le groupe 2, z = 2.11, p <.05). De nouveau, on retrouve une certaine hiérarchie entre niveau scolaire et coût cognitif dans la réalisation des épreuves. Comme pour les épreuves testant l'étape visuelle les performances du groupe 2 sont plus proches de celles du groupe 3 (PHO : +79 ms; PH2 : +164) que de celles du groupe 1 (respectivement +199 et +282 ms), excepté pour PH1 où la différence est dans les deux cas de 230 ms. On peut également relever une particularité du groupe 3 :

il se différencie des deux autres groupes dans la mesure où l'écart de 39 ms observé entre PH1 et PH2 s'avère non significatif (z =.40, p <.34). Le fait d'être face à un mot et un non-mot plutôt que face à deux non-mots ne modifie donc en rien les performances des sujets du groupe 3, comme si ce groupe procédait de la même façon dans les deux cas.

Les résultats mettent en évidence des difficultés au niveau de l'étape de traitement phonologique avec un déficit plus marqué dans le cas des lecteurs du groupe 3.

3.3.3 *Étape lexicale* [2]

Quelles sont les différences observables entre les trois groupes du point de vue de l'activation des mots ?

Le Tableau 6 présente les résultats pour quatre épreuves testant l'activation des entrées lexicales par la voie directe et des concepts au sein du lexique.

TABLEAU 6

Coût moyen (en ms) pour les épreuves testant l'étape lexicale pour chacun des trois groupes

	Groupes			
	1	2	3	Moy
DL1	185 (54)	286 (117)	378 (249)	272 (162)
DL2	184 (46)	254 (103)	323 (182)	245 (124)
DL3	98 (36)	127 (63)	181 (187)	130 (106)
SEM	399 (101)	545 (193)	660 (311)	519 (225)

Note : écarts-types entre parenthèses.
Décision lexicale 1 (DL1), 2 (DL2) et 3 (DL3) et catégorisation sémantique (SEM).

L'analyse de variance sur les trois épreuves de décision lexicale met en évidence un effet du groupe (F (2, 155) = 35.80, p <.001) : quelle que soit l'épreuve, le groupe 1 est plus rapide (cf. Annexe 2); un effet de la tâche (F (2, 310) = 252.17, p <.001) ainsi qu'une interaction entre groupe et tâche (F (4, 310) = 10.59, p <.001). La décision lexicale 3 est réussie plus rapidement que la décision lexicale 2 (groupe 1 : z = 11.12, p <.01; groupe 2 : z = 8.43, p <.01; groupe 3 : z = 3.33, p <.01) et que la décision lexicale 1 (respec-

2 Nous ne présenterons pas ici les résultats de l'épreuve d'amorçage orthographique, la moyenne des temps de réaction (moyennant des effets de facilitation et d'inhibition) n'étant pas pertinente.

tivement, $z = 10.12$, $z = 9.64$ et $z = 3.90$, $p <.01$). La plus grande rapidité en DL3 peut sans doute s'expliquer par le fait que la fréquence des mots utilisés est plus élevée dans cette épreuve que dans les deux autres épreuves de décision lexicale. En revanche, seul le groupe 2 semble présenter un coût légèrement différent entre DL1 et DL2 (groupe 2, $z = 1.65$, $p <.05$; groupe 1, $z =.11$, $p <.46$; groupe 3, $z = 1.06$, $p <.14$).

Enfin, les résultats pour la catégorisation sémantique (SEM) font apparaître que le coût diminue quand le niveau scolaire augmente (cf. Annexe 2). Il n'y a rien ici de très surprenant dans la mesure où, pour activer la signification d'un mot écrit, il faut préalablement activer une entrée lexicale et que l'analyse des résultats montre que les processus impliqués dans cette activation sont déficients.

De nouveau, les épreuves montrent que les lecteurs les plus lents sont plus lents dans l'activation des entrées lexicales. Toutefois, dans la mesure où il existe une certaine hiérarchie entre les différentes étapes de traitement, on ne peut pas parler, à ce niveau de l'analyse, d'une difficulté spécifique, les lecteurs ayant dû effectuer auparavant un certain nombres d'opérations (au moins visuelles, si ce n'est phonologiques).

Cette première analyse des résultats a permis de mettre en évidence les différences existant entre les trois groupes. En moyenne, les groupes 2 et 3 se différencient systématiquement du groupe 1 par des coûts supérieurs. Comme en lecture, le groupe 2 présente des résultats intermédiaires. Cependant, contrairement à la vitesse de lecture, le groupe 2 est souvent plus proche des résultats du groupe 3 que de ceux du groupe 1; et pour deux épreuves (COD, PHO), les différences observées avec le groupe 3 ne sont statistiquement pas significatives. Cette progression, dans les coûts associés aux épreuves, au travers des trois groupes, peut sans doute être mise en relation avec l'hypothèse qu'il existerait un continuum dans l'automatisation des processus. D'ailleurs, l'étude des coefficients de corrélation entre les vitesses moyennes de lecture (en lecture de textes -LMM- et en lecture de listes de mots -LHC) et les coûts associés à chaque épreuve d'évaluation (cf. Tableau 7) montre qu'aucune corrélation significative n'est observée entre les épreuves de lecture et les 12 épreuves testant les processus impliqués dans la reconnaissance des mots pour le groupe 1; alors que 9 épreuves pour le groupe 2 et 10 épreuves pour le groupe 3 sont corrélées aux deux épreuves de lecture.

Tout semble donc se passer comme si en lecture, pour les groupes 2 et 3, une partie importante des ressources attentionnelles était allouée aux processus menant à la reconnaissance des mots.

En ce qui concerne les relations entre les processus qui mènent à la reconnaissance des mots, l'étude des coefficients de corrélation fait ressortir la complexité des liens qu'entretiennent les différentes épreuves en fonction du niveau de lecture : on pouvait s'attendre, d'une part, à ce que les épreuves regroupées au sein d'une même étape soient à la fois plus souvent et plus fortement intercorrélées entre elles qu'avec les épreuves testant d'autres étapes;

d'autre part, à ce qu'il n'y ait pas ou peu de différences entre les groupes. Ce n'est pas le cas (cf. Annexe 3). Pour les groupes 1 et 2, le nombre important de corrélations entre les épreuves de diagnostic (de 8 à 10 épreuves corrélées pour le groupe 1; de 9 à 10 pour le groupe 2) suggère que ces épreuves mettent en oeuvre des processus qui, s'ils ne sont pas identiques, sont suffisamment proches (ou nécessitent les mêmes ressources attentionnelles) pour qu'une analyse des corrélations ne permette pas de les dissocier. Le groupe 3 apparaît différent dans la mise en oeuvre des processus. En effet, les processus semblent beaucoup moins dépendants les uns des autres pour ce groupe (de 0 à 8 épreuves sont corrélées) que pour les groupes 1 et 2. On peut se demander dans quelle mesure il n'existerait pas des systèmes de traitement de l'information différents selon le niveau d'efficience en lecture (peut-être par la mise en oeuvre de systèmes compensatoires ?). En attendant de pouvoir répondre à cette question, voyons quelles sont les différences observables au sein d'un même groupe.

TABLEAU 7

Coefficients de corrélation entre les vitesses en lecture de textes (LMM) et en lecture de listes de mots (LHC) et les coûts moyens dans les différentes épreuves linguistiques pour chacun des trois groupes.

	Groupes					
	1		2		3	
	LMM	LHC	LMM	LHC	LMM	LHC
COD	.20	.11	.12	.16	.19	.04
JIL	.12	20	.38**	.43**	.61**	.57**
SM1	.19	.20	.34**	.32*	.60**	.51**
SM2	.14	.08	.52**	.49**	.68**	.71**
PHO	.11	.13	.32**	.25*	-.17	-.30
PH1	.19	.23	.44**	.43**	.63**	.70**
PH2	.17	.09	.41**	.37**	.62**	.64**
DL1	.19	.16	.38**	.32**	.50**	.60**
DL2	.09	-.01	.27*	.27*	.57**	.59**
DL3	.16	.10	.10	.09	.43**	.50**
SEM	.25	.11	.20	.12	.49**	.58**

Note : * p<.05, ** p<.01. Quand les sujets commettaient en moyenne plus de 50% d'erreurs dans une épreuve, nous n'avons pas pris en compte leur temps de réaction. Nous précisons ici le nombre de sujets lorsque celui-ci est inférieur à celui attendu. SM1: Groupe 2 (effectif normal: 65 sujets): 61 sujets. Groupe 3 (effectif normal: 38 sujets): 37 sujets. PHO: Groupe 2, 60 sujets; Groupe 3, 30 sujets. PH1: Groupe 2, 64 sujets; Groupe 3, 36 sujets. PH2: Groupe 2, 63 sujets; Groupe 3, 32 sujets. DL2: Groupe 2, 64 sujets; Groupe 3, 37 sujets. SEM: Groupe 3, 37 sujets.

3.3.4 *Quelles différences peut-on observer entre les «mauvais» lecteurs ? Analyse des profils*

Nous avons vu au cours de l'analyse précédente que les épreuves permettaient de mettre en évidence des différences entre les trois groupes de ni-

veau du point de vue du coût associé à la mise en oeuvre des différents processus impliqués dans la reconnaissance des mots. L'objectif de cette partie sera de montrer que ces épreuves permettent de mettre en évidence des difficultés différentes d'un «mauvais» lecteur à l'autre. Pour ce faire nous avons analysé les distributions effectuées lors de l'étalonnage.

Pour l'ensemble de la population (160 sujets), une distribution en 9 classes a été effectuée pour chaque coût après transformation des mesures (d'abord en logarithme puis en notes centrées réduites). De cette façon, il est possible de situer (pour chaque indicateur) chacun des sujets dans une des 9 classes ainsi établies (Zagar, Jourdain et Lété, 1993). A partir du positionnement de chaque sujet, nous avons repris les distributions en séparant chacun des groupes (cf. pour exemple les Figures 2 et 3). Pour effectuer une analyse des difficultés spécifiques à chaque sujet, nous avons pris uniquement en compte les sujets des groupes 2 et 3 qui se situaient au-delà de la distribution des sujets du groupe 1. A deux exceptions près (en SM1, le sujet 149 et en PH2 le sujet 105), on ne trouve pas de sujets du groupe 1 au-delà de la sixième classe, qui correspond à la moyenne générale plus 1.125 écart-type.

Comment les sujets des groupes 2 et 3 se répartissent-ils dans chacune des épreuves ? (cf. Annexes 4 à 6)

En ce qui concerne les épreuves de lecture, 33 % des sujets obtiennent des temps supérieurs pour au moins l'une des deux épreuves de lecture. Dans la majorité des cas (59 %) les deux vitesses de lecture sont inférieures (i.e. au-delà de 924 ms/mot en LMM et de 1191 ms/mot en LHC); dans 12 % des cas seule la lecture en contexte est plus lente (LMM) et dans 29 % des cas seule la lecture de listes de mots est plus lente (LHC) (cf. Tableau 8 pour la répartition en fonction du niveau scolaire).

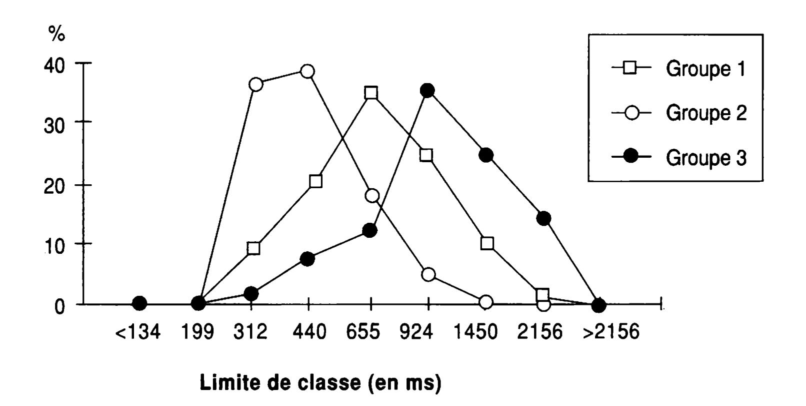


FIGURE 2

Distribution des vitesses moyennes de lecture en LMM selon le niveau scolaire

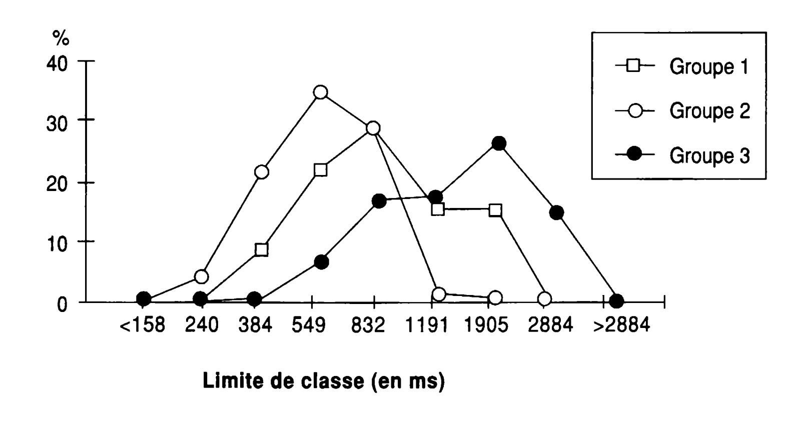


FIGURE 3

Distribution des vitesses moyennes de lecture en LHC selon le niveau scolaire

Qu'en est-il des épreuves de diagnostic ?

Sur l'ensemble des sujets qui obtiennent une vitesse de lecture inférieure à celle du groupe 1, seulement deux sujets (appartenant au groupe 2) ne présentent aucune difficulté particulière du point de vue du coût associé aux épreuves de diagnostic. Dans 94 % des cas, un temps de lecture supérieur est associé à au moins une difficulté au niveau des processus menant à la reconnaissance des mots.

Au total 31 sujets du groupe 2 (47%) et 33 sujets du groupe 3 (87%) présentent une difficulté dans au moins une épreuve (cf. Annexes 4 à 6). En revanche, dans cette population, respectivement 64 et 39% des sujets obtiennent des temps de lecture équivalents aux bons lecteurs (groupe 1). Si, comme nous l'avons vu précédemment, la vitesse de lecture peut être presque systématiquement associée à une difficulté au niveau des processus impliqués dans la reconnaissance des mots, en revanche dans 51% des cas elle ne permet pas de détecter une difficulté au niveau de ces processus (cf. Tableau 9).

Le Tableau 8 donne, par épreuve et pour chaque groupe, le pourcentage de sujets qui se situent au-delà de la distribution en 9 classes des coûts du groupe 1.

Sur l'ensemble des épreuves entre 5 et 30 pour-cent des sujets des groupes 2 et 3 (103 sujets) présentent au moins une difficulté, SM1 et PHO étant les épreuves pour lesquelles les sujets présentent un maximum de difficultés (29 et 28% des sujets sont au-delà du groupe 1); les épreuves de décision lexicale (DL1, 2 et 3) étant celles qui présentent le moins de difficultés.

Cependant, si nous nous intéressons à ce qui se passe pour chaque groupe, nous pouvons noter que ce sont majoritairement les sujets du groupe 3 qui sont en difficulté. Pour ce groupe, l'ensemble des épreuves (à l'exception de DL3) apparaît concerné. En effet, entre 27 et 54 % des sujets sont au-delà

des coûts observés pour le groupe 1. Les problèmes les plus importants se situant au niveau des épreuves testant la phonologie (cf. Tableau 8).

TABLEAU 8

Par épreuve, pourcentage de sujets des groupes 2 et 3 (N = 103) dont le coût moyen est supérieur à A

	A (ms)	Groupes 2	3	Total
Effectif	**65**	**38**	**103**	
COD	249	12	29	18
JIL	505	14	34	21
SM1	1835	21 (26)	40	29 (31)
SM2	942	15	39	24
PHO	731	20 (25)	43 (54)	28 (36)
PH1	1010	12 (14)	44 (47)	24 (26)
PH2	1081	13 (15)	31 (42)	19 (25)
DL1	455	8	16	11
DL2	385	9 (11)	27 (29)	16 (13)
DL3	249	3	8	5
SEM	698	15	27 (29)	20 (20)
LMM	924	1	8	4
LHC	1191	8	13	10
LMM & LHC		9	37	19

Note : critère de difficulté, égal à la moyenne générale + 1.125 écart-type; entre parenthèses pourcentage intégrant les sujets commettant plus de 50% d'erreurs.

Pour le groupe 2, les sujets en difficultés représentent au maximum un quart de cette population (si on tient compte des sujets pour qui la difficulté se manifeste sur le pourcentage d'erreurs), ces difficultés concernant principalement SM1 et PHO.

Quels sont les profils des sujets présentant une difficulté ?

Pour chaque sujet présentant une difficulté (i.e. au-delà du groupe 1), nous avons effectué un diagnostic en terme de difficulté(s) visuelle et/ou phonologique et/ou lexicale (cf. Annexes 4 à 6). Le Tableau 9 synthétise ces résultats en dissociant les profils associés ou non à une vitesse de lecture plus lente que celle du groupe 1.

Nous pouvons différencier 3 groupes principaux. Le premier concerne les sujets qui présentent des difficultés à un seul niveau soit visuel (25 %), soit phonologique (9 %), soit lexical (7 %).

Quand la difficulté se situe exclusivement au niveau des processus mis en oeuvre dans l'étape visuelle, elle peut concerner :

a) l'activation des codes graphémiques (comme c'est le cas pour les sujets 203, 307 et 355; cf. Annexe 4),

b) l'analyse d'une chaîne de caractères (sujet 330) et/ou la comparaison de deux chaînes de caractères (sujets 228, 337, 208, 257, 239, 251, 261, 322, 334, 336, 319),

c) les deux (sujet 207).

Si la difficulté se manifeste uniquement pour l'étape phonologique, elle est essentiellement reflétée par la décision lexicale phonologique (cf. Annexe 4). Autrement dit le problème se situe au niveau de l'activation d'une entrée lexicale par le biais des codes phonémiques. On peut également remarquer qu'une difficulté à ce seul niveau ne donne pas lieu à un ralentissement de la vitesse de lecture.

Quand la difficulté est présente pour la seule étape lexicale, on peut différencier les problèmes au niveau de l'activation des unités lexicales (sujet 236), des problèmes au niveau de l'activation des concepts (sujets 205 et 227).

TABLEAU 9

Répartition (en %) des 64 sujets présentant une ou plusieurs difficultés au niveau des processus impliqués dans la reconnaissance des mots en fonction de la vitesse de lecture

	Vitesse de lecture		
Effectif	**inférieure**	**supérieure**	**TOTAL**
VIS	14	11	25
PHO	8	0	8
LEX	5	2	7
VIS + PHO	12	9	22
VIS + LEX	3	0	3
LEX + PHO	5	6	11
VIS + PHO + LEX	5	21	25
TOTAL	**51**	**49**	**100**

Le second groupe concerne les sujets qui présentent des difficultés à deux niveaux : visuel et lexical (3 %), visuel et phonologique (22 %), phonologique et lexical (11%). Les sujets réunis dans ces trois sous-groupes présentent des profils très différents (annexe 5). Pour les difficultés visuelles et lexicales, on retrouve la diversité des problèmes énumérés précédemment dans le cas d'une difficulté à un seul niveau (visuel ou lexical). Par contre en ce qui concerne les difficultés phonologiques, on voit apparaître des difficultés au seul niveau de l'activation des codes phonémiques (sujets 311, 320, 303), ou combinées avec des problèmes dans l'activation des unités lexicales par le biais de ces codes phonémiques (sujets 333, 305, 315,....).

Enfin, le dernier groupe (cf. Annexe 6) concerne les sujets qui ont des difficultés à tous les niveaux (25 %). Là encore, il existe une diversité au sein

de ce groupe entre des personnes qui présentent des difficultés massives dès l'activation des codes graphémiques (par exemple le sujet 327) et des personnes qui présentent des difficultés qui semblent moins importantes (par exemple le sujet 233).

A partir de ces premières analyses, il apparaît possible de mettre en évidence des difficultés différentes d'un «mauvais» lecteur à l'autre; il convient toutefois de préciser que les analyses rapportées ici n'ont pris en compte que les coûts moyens pour les épreuves. Les profils pourront donc être affinés par la prise en compte d'autres informations comme le taux d'erreurs et les effets associés aux variables manipulées dans chaque épreuve.

Le taux d'erreurs peut venir confirmer une difficulté observée à partir des mesures chronométriques, en signalant que la mise en oeuvre des processus est non seulement coûteuse mais aussi inefficace dans la mesure où elle conduit fréquemment à un échec. Le taux d'erreurs peut également mettre en évidence une stratégie non observable à partir des temps de réaction, le sujet pouvant privilégier la rapidité des réponses au dépend de leur précision (speed-accuracy trade-off).

En ce qui concerne les variables manipulées, l'étude des effets observés en fonction des caractéristiques du matériel devrait permettre d'approfondir et de spécifier l'origine des difficultés. Par exemple, un effet de longueur en décision lexicale permet de penser que le sujet effectue l'analyse des mots de façon sérielle ce qui n'est pas le cas pour le lecteur habile.

En apportant de nouvelles techniques, de nouvelles conceptions, la psychologie cognitive est en mesure d'aider le formateur (en formation initiale ou continue) à mieux appréhender les problèmes qui peuvent se poser à tel ou tel apprenant au cours de sa formation. Cependant l'utilisation d'outils d'évaluation existant ou la construction d'outils de remédiation adaptés aux difficultés observées passe, pour être pleinement satisfaisante, par un investissement important et une appropriation de la démarche en psychologie cognitive.

Annexe 1

Exemples de matériel utilisé dans chacune des épreuves en fonction des étapes testées

1. Étape de traitement visuel

1.1. Codage abstrait des lettres (COD)

[Stimulus]	A z	A a	L L
[Consigne]	"Les 2 lettres sont-elles identiques ?"		
[Réponse]	NON	OUI	OUI

1.2. Jugement d'identité de lettres (JIL)

[S]	AAA	ABA	AAAAAAAAAAB
[C]	"Toutes les lettres sont-elles identiques ?"		
[R]	OUI	NON	NON

1.3. Jugement de similitude : suite de lettres non-prononçable (SM1)

[S]	ACFDX ACFDX	BCDTF BCDTG
[C]	"Les 2 suites sont-elles identiques ?"	
[R]	OUI	NON

1.4. Jugement de similitude: suite de lettres prononçable (SM2)

[S]	ACFDX ACFDX	BCDTF BCDTG
[C]	"Les 2 suites sont-elles identiques ?"	
[R]	OUI	NON

2. Étape de traitement phonologique

2.1. Décision phonologique ‡ choix forcé avec deux non-mots (PH1)

[S]	aumir eaumir	romau romou
[C]	"Les deux items se prononcent-ils de la même manière?"	
[R]	OUI	NON

2.2. Décision phonologique ‡ choix forcé avec un mot et un non-mot (PH2)

[S]	aumir eaumir	romau romou
[C]	"Les deux items se prononcent-ils de la même manière?"	
[R]	OUI	NON

2.3. Décision lexicale phonologique (PHO)

[S]	phoule	romau
[C]	"Le stimulus se prononce-t-il comme un mot de la langue ?"	
[R]	OUI	NON

3. Étape de traitement lexical

3.1. Décision lexicale (DL1, DL2, DL3)

[S]	bijou	carrosse	typrfg	trapoue
[C]	"Le stimulus est-il un mot de la langue ?"			
[R]	OUI	OUI	NON	NON

3.2. Amorçage orthographique (AOR)

[A]	CANARD	-----	natte
[S]	canard	étain	talin
[C]	"Le stimulus est-il un mot de la langue ?"		
[R]	OUI	OUI	NON

3.3. Catégorisation sémantique (SEM)

[S]	ANIMAL chien	ANIMAL pieuvre	POISSON baleine	POISSON chien
[C]	"Le second item appartient-il ‡ la même catégorie que le premier ?"			
[R]	OUI	OUI	NON	NON

Annexe 2

Comparaison de moyennes pour les coûts associés aux épreuves testant les étapes visuelle, phonologique et lexicale et pour les vitesses de lecture.

Le test z $[z = (m_1 - m_2) / \sqrt{(s_1^2/n_1) + (s_2^2/n_2)}]$ permet de comparer des moyennes dans le cas de grands échantillons (>30 sujets).

Comparaison	Groupe 1 vs Groupe 2	Groupe2 vs Groupe 3
COD	z = 2.91*	z = .83 ns
JIL	z = 5.57**	z = 2.11*
SM1	z = 3.95**	z = .2.45*
SM2	z = 5.74**	z = 2.44*
PHO	z = 6.29**	z = 1.41 ns
PH1	z = 6.03**	z = 3.30**
PH2	z = 6.57**	z = 1.96*
DL1	z = 6.24**	z = 2.14*
DL2	z = 4.88**	z = 2.12*
DL3	z = 3.16**	z = 1.72*
SEM	z = 5.32**	z = 2.04*
LMM	z = 5.98**	z = 4.29**
LHC	z = 6.02**	z = 4.83**

* p < .05 (z = 1.645)
** p < .01 (z = 2.325)
ns: non significatif

Annexe 3

Matrices de corrélations entre les coûts associés aux différentes épreuves pour les trois groupes.

Groupe 1

	COD	JIL	SM1	SM2	PHO	PH1	PH2	DL1	DL2	DL3
COD	1									
JIL	.34*	1								
SM1	.09	.45**	1							
SM2	.36**	.56**	.59**	1						
PHO	.24	.43**	.44**	.67**	1					
PH1	.35**	.68**	.54**	.73**	.65**	1				
PH2	.26*	.49**	.54**	.80**	.68**	.71**	1			
DL1	.49**	.43**	.30*	.54**	.56**	.58**	.56**	1		
DL2	.52**	.41**	.23	.61**	.54**	.44**	.66**	.61**	1	
DL3	.25*	.45**	.27*	.45**	.59**	.42**	.48**	.54**	.46**	1
SEM	.21	.36**	.39**	.54**	.62**	.59**	.69**	.50**	.40**	.45**

Groupe 2

	COD	JIL	SM1	SM2	PHO	PH1	PH2	DL1	DL2	DL3
COD	1									
JIL	.49*	1								
SM1	.18	.53**	1							
SM2	.46**	.70**	.49**	1						
PHO	.11	.32*	.44**	.47**	1					
PH1	.31*	.60**	.58**	.80**	.57**	1				
PH2	.33**	.53**	.55**	.71**	.82**	.83**	1			
DL1	.50**	.47**	.32*	.70**	.61**	.66**	.63**	1		
DL2	.46**	.49**	.32*	.58**	.53**	.55**	.71**	.76**	1	
DL3	.44*	.42**	.25*	.31*	.43**	.33**	.51**	.55**	.62**	1
SEM	.25*	.29*	.23	.35**	.72**	.41**	.60**	.66**	.55**	.41**

Groupe 3

	COD	**JIL**	**SM1**	**SM2**	**PHO**	**PH1**	**PH2**	**DL1**	**DL2**	**DL3**
COD	1									
JIL	.36*	1								
SM1	.04	.56**	1							
SM2	.20	.66**	.70**	1						
PHO	.26	-.19	-.10	-.14	1					
PH1	.11	.57**	.76**	.86**	-.07	1				
PH2	.05	.33	.72**	.78**	-.08	.86**	1			
DL1	.01	.14	.36*	.69**	-.14	.73**	.76**	1		
DL2	.12	.31	.45**	.69**	-.11	.61**	.72**	.80**	1	
DL3	-.01	.11	.33*	.55**	-.06	.73**	.77**	.87**	.65**	1
SEM	.05	.24	.41*	.70**	-.26	.66**	.82**	.81**	.65**	.74**

Note :*p<.05, **p<.01

Annexe 4

Sujets présentant une difficulté au niveau d'une seule étape

Sujets	COD	JIL	SM1	SM2	PHO	PH1	PH2	DL1	DL2	DL3	SEM	LMM	LHC	Diff
203	+													V
307	+											+		V
355	+													V
207	+		+									+	+	V
228		+	+										+	V
337		+	+										+	V
208			+	+								+	+	V
330		+												V
257			er											V
239			+											V
251			+											V
261			+											V
322			+											V
334			+									+	+	V
336			+									+	+	V
319				+										V
236									+					L
205											+	+		L
227											+			L
230					er									P
264					+									P
312					+									P
321					+									P
309					er									P

Note : V = difficulté(s) visuelle(s), P = difficulté(s) phonologique(s), L = difficulté(s) lexicale(s)
+ = difficulté(s) sur le coût de l'épreuve, er = difficulté(s) sur l'exactitude

Annexe 5

Sujets présentant une difficulté au niveau de deux étapes

Sujets	COD	JIL	SM1	SM2	PHO	PH1	PH2	DL1	DL2	DL3	SEM	LMM	LHC	Diff
203	+													V
307	+											+		V
355	+													V
207	+		+									+	+	V
228		+	+										+	V
337		+	+										+	V
208			+	+								+	+	V
330		+												V
257			er											V
239			+											V
251			+											V
261			+											V
322			+											V
334			+									+	+	V
336			+									+	+	V
319				+										V
236									+					L
205											+	+		L
227											+			L
230					er									P
264					+									P
312					+									P
321					+									P
309					er									P

Note : V = difficulté(s) visuelle(s), P = difficulté(s) phonologique(s), L = difficulté(s) lexicale(s)
+ = difficulté(s) sur le coût de l'épreuve, er = difficulté(s) sur l'exactitude

Annexe 6

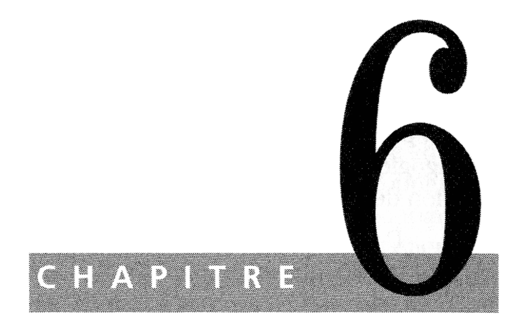
Sujets présentant une difficulté au niveau de trois étape

Sujets	COD	JIL	SM1	SM2	PHO	PH1	PH2	DL1	DL2	DL3	SEM	LMM	LHC	Diff *
327	+	+	+	+	er	+	er	+	+	+	er	+	+	VPL
214	+	+	+	+	er	+	+	+	+	+			+	VPL
338	+	+	+	+	er	+	er	+	+		+	+	+	VPL
326	+	+	+	+	+	+	+		+		+	+	+	VPL
335	+	+	+	+	+	+	+		+		+	+	+	VPL
232	+	+	+	+	+		+	+	+		+		+	VPL
250	+	+	er	+	+	+	+				+	+		VPL
310	+		+	+	+	+	+	+			+			VPL
209	+	+		+	er	er	er	+			+	+	+	VPL
331		+	+	+	er	+	er	+	+	+	+	+	+	VPL
302			+	+	+	+	+	+	er	+	+	+	+	VPL
202			+	+	+	+	+	+	er					VPL
245			+		+	+	+				+	+	+	VPL
329			+	+	+		+	+	+		+	+	+	VPL
306			+		+	+	+		+			+	+	VPL
233			+		+						+			VPL

Sujets ne présentant aucune difficulté au niveau des trois étapes

Sujets	COD	JIL	SM1	SM2	PHO	PH1	PH2	DL1	DL2	DL3	SEM	LMM	LHC	Diff *
221												+	+	
243												+	+	

Note : VPL = difficulté(s) visuelle(s), phonologique(s) et lexicale(s)
+ = difficulté(s) sur le coût de l'épreuve, er = difficulté(s) sur l'exactitude



La remédiation des difficultés de lecture par la rétroaction verbale sur ordinateur : bilan des recherches

Bernard Lété

Introduction

«Faire parler» l'ordinateur pour améliorer la lecture chez l'enfant qui apprend à lire constitue aujourd'hui un champ particulièrement novateur d'application des recherches de psychologie cognitive. Cette nouvelle démarche repose en fait sur une réalité solidement ancrée dans les pratiques pédagogiques : le rôle de la communication orale comme support de la transmission des savoirs.

De l'ordinateur simplement calculateur, nous sommes passés à l'ère du multimédia où l'ordinateur est à la fois lecteur (reconnaissance de caractères), interlocuteur (production de la parole) et auditeur (reconnaissance de la parole). Cette véritable mutation technologique que nous vivons produit elle-même une double mutation. Mutation pédagogique d'abord, où la psychologie cognitive tient un rôle majeur en fournissant aux professionnels de l'enseignement et de la formation des modèles d'apprentissage et d'expertise que ceux-ci doivent s'approprier. Mutation sociale également car de nouveaux comportements d'apprentissage voient le jour. Leong (1992b) affirme que ces nouveaux comportements «restructurent *notre fonctionnement mental et réorganisent des habiletés cognitives de haut niveau*» (p.95).

De nombreux chercheurs, travaillant dans le domaine de l'enseignement assisté par ordinateur, invitent les enseignants à faire utiliser les ordinateurs par les enfants non pas comme un outil de travail mais comme un objet de pensée. Nous ne sommes pas loin des conceptions de Piaget : l'ordinateur, dans les classes, peut permettre à l'enfant d'être le bâtisseur de ses propres structures intellectuelles grâce à la médiatisation qu'il peut exercer entre le langage et le développement cognitif.

L'enjeu de la mise au point de systèmes d'aide à la lecture assisté par ordinateur est considérable. Comme le souligne Haton (1992), il s'agit en effet de faire bénéficier l'enseignement des nouvelles techniques informatiques tout en brisant le rôle central et dirigiste de la machine et la rigidité de ses échanges avec l'apprenant. Il s'agit d'introduire de la souplesse, favoriser une progression et même une démarche individualisée, sanctionner à bon escient. En un mot, il faut repenser l'image de «l'ordinateur tuteur» à travers les nouveaux moyens que le multimédia nous offre [1].

Le présent travail se limite à un objectif précis. Il présente les résultats de recherches (essentiellement anglo-saxonnes) qui ont créées et testées des systèmes de rétroaction verbale pour améliorer, d'une façon générale, les habiletés de reconnaissance des mots chez des enfants présentant, à des degrés plus ou moins sévères, des difficultés d'apprentissage. Ces systèmes, pour la plupart, utilisent la parole de synthèse comme rétroaction sonore [2]. Bien que

1 Voir de La Passardière et Baron (1991) pour des exemples d'applications de ces technologies nouvelles.

2 Pour une perspective plus didactique, le lecteur pourra utilement se reporter à l'ouvrage de Cohen (1992) qui fait part de plusieurs travaux français sur l'utilisation de la synthèse vocale dans l'apprentissage et le perfectionnement de la langue écrite.

les recherches soient encore très restreintes et les résultats tangibles peu nombreux, nous considérons qu'il est utile d'établir un premier bilan car nous pensons que ce champ d'étude est fructueux. Deux raisons ont guidé notre choix.

La première est liée à la place de plus en plus importante que tient la psychologie cognitive dans le domaine de l'éducation : elle a apporté en effet de puissants concepts et modélisations de l'acquisition et du développement du savoir. On assiste peut-être, depuis les années 1980, au développement d'une nouvelle discipline qui tend à émerger essentiellement des recherches sur la lecture. Cette nouvelle discipline, que l'on pourrait nommer les *sciences cognitives appliquées à l'éducation*, contribue en particulier à l'élaboration d'outils d'évaluation et de remédiation capables d'aider les enseignants et les formateurs dans leur tâche[3]. Les systèmes d'aide aux apprentissages se sont en effet considérablement développés; aujourd'hui, certains systèmes, dits experts, utilisent des techniques très pointues de modélisation des connaissances et de stockage des actions des utilisateurs. La rétroaction verbale est à l'heure actuelle une sophistication particulière de ces systèmes d'aide; elle s'inscrit donc dans un champ déjà très vaste et particulièrement important pour les professionnels de l'enseignement.

La seconde raison est liée à la difficulté des chercheurs à transférer les résultats de leurs recherches dans le domaine de l'enseignement. Il faut reconnaître qu'il existe encore des «résistances» quant à l'utilisation de tels systèmes (en particulier ceux qui concernent l'évaluation et la remédiation des capacités cognitives). Pour utiliser un terme qui fait souvent partie du vocabulaire des détracteurs de ces outils (détracteurs qui, bien souvent, ne sont pas directement impliqués dans le difficile problème de la transmission des savoirs), le domaine de l'éducation doit *faire le deuil* de plusieurs concepts archaïques qui ont longtemps prévalus pour expliquer de façon exclusive les difficultés d'apprentissage -en particulier l'importance des facteurs dits socio-affectifs-; concepts dont l'apparente richesse sémantique masque bien souvent le caractère superficiel et fallacieux des explications avancées. Les résultats des travaux de psychologie cognitive nous conduisent à centrer notre curiosité sur les mécanismes d'acquisition et de transmission du savoir tout en gardant à l'esprit le rôle important, d'une part, des représentations mentales, qu'il convient d'analyser en dehors de conceptions neuro-biologique ou socio-culturelle; et, d'autre part, de la psychologie dite computationnelle et des ordinateurs dans la modélisation de la pensée humaine[4].

Les objectifs de ce travail s'inscrivent donc dans cette logique. A travers le thème abordé, il s'agit pour nous d'illustrer comment la psychologie co-

---

3 Cfr. pour le problème de l'illettrisme, le logiciel d'Evaluation-Diagnostic des Capacités Cognitives en Lecture pour Adultes (*ECCLA*; Zagar, Jourdain & Lété, 1993).

4 Voir Leong (1993) pour une discussion du rôle de la psychologie cognitive dans le domaine de l'éducation et Gardner (1985) pour un point de vue plus général sur le développement de la psychologie cognitive.

gnitive a profondément modifié la façon d'appréhender les difficultés d'apprentissage (ici les difficultés de lecture). En s'appuyant sur des modèles issus de la recherche fondamentale, les chercheurs ont pu élaborer des outils d'évaluation et de remédiation qui ne sont pas a-théoriques, comme cela est encore trop souvent le cas aujourd'hui.

Après avoir montré l'importance de l'efficacité des traitements phonologiques dans l'apprentissage de la lecture et décrit quels peuvent être les effets de la rétroaction verbale sur leurs déficits, nous décrivons quelques aspects techniques de la production de la parole sur ordinateur. Etant donné que la plupart des systèmes, dans les recherches présentées ici, utilisent de la parole de synthèse, nous abordons le problème de sa compréhension. Nous présentons alors les recherches les plus récentes, essentiellement en langue anglaise, qui ont été menées, d'une part, pour améliorer le décodage et la reconnaissance des mots et, d'autre part, pour améliorer la compréhension.

1. LA RÉTROACTION VERBALE ET SES EFFETS SUR LES TRAITEMENTS PHONOLOGIQUES

De nombreux travaux rendent compte de l'importance des traitements phonologiques pour apprendre à lire dans un système alphabétique. Wagner et Torgesen (1987) recensent trois habiletés liées au traitement phonologique de la langue susceptibles d'avoir un rôle causal dans l'acquisition de la lecture.

La première, la conscience phonologique, est la capacité à identifier les composants phonologiques des unités linguistiques et à les manipuler de façon intentionnelle (cf. Gombert, 1990[5]). Elle est généralement appréhendée dans des tâches de segmentation de mots en phonèmes, en syllabes ou en unités subsyllabiques (attaque-rime[6] par exemple). Avec l'apprentissage de la lecture, la performance des enfants à ce type de tâche évolue très rapidement : si les lecteurs débutants échouent à segmenter les mots en phonèmes, ils réussissent par contre à la segmentation en syllabes (Wise, Olson & Treiman, 1990).

La seconde, le décodage, est la capacité à relier les unités du langage parlé (les phonèmes) aux unités du langage écrit (les graphèmes) pour accéder au lexique mental. Le rôle du décodage dans les difficultés de lecture reste controversé; il apparaît cependant que, lors de l'accès au lexique, il est utilisé par tous les lecteurs en fonction de la familiarité du mot lu et qu'il est très important dans les premières étapes de l'acquisition de la lecture. Pour Olofsson (1992), les problèmes du décodage ont une incidence à la fois sur la compré-

5 Gombert (1990) utilise le terme de «*capacité métaphonologique*».

6 La syllabe peut être décomposée en deux parties : l'attaque qui correspond au groupe de consonne(s) initiale(s) et la rime qui correspond au(x) phonème(s) qui suivent. Par exemple, l'attaque du mot *front* est /fr/, la rime étant /ont/.

hension et sur la pratique de la lecture. Si le décodage est déficient, l'information est traitée par d'autres moyens, en particulier par une utilisation anormale du contexte. Le fait que les enfants mauvais compreneurs ont invariablement une habileté médiocre dans la reconnaissance des mots mais montrent d'importants effets du contexte linguistique dans de nombreuses tâches a longtemps été un paradoxe relevé dans la littérature sur les difficultés de lecture; paradoxe qui était lui-même en contradiction avec certaines théories développementales (Smith, 1971). Il a été expliqué par Stanovich (1980) de la façon suivante : le manque d'efficacité des processus de reconnaissance des mots chez les mauvais lecteurs les force à mobiliser leurs ressources attentionnelles sur des mécanismes contextuels ce qui diminue leurs ressources disponibles pour des processus de plus haut niveau (la compréhension). A l'opposé, l'efficacité des mécanismes de reconnaissance des mots des bons lecteurs ne nécessite pas de ressources attentionnelles, celles-ci peuvent dès lors être déléguées vers la compréhension. Les difficultés du décodage provoquent également un désintérêt pour la lecture. La probabilité de rencontrer des mots nouveaux diminue limitant ainsi le développement du vocabulaire. Le désintérêt engendre lui-même moins de pratique, c'est à dire moins d'expérience avec l'écrit, alors que celle-ci est fondamentale à l'acquisition d'un lexique orthographique qui permet une reconnaissance rapide et efficace des mots (Cunningham & Stanovich, 1990; Stanovich & West, 1989).

La dernière enfin, le recodage phonétique en mémoire de travail, est la capacité à maintenir activés les sons précédemment décodés de façon à les assembler pour former un mot. Un recodage efficace de l'information phonologique en mémoire de travail permet au lecteur d'allouer un maximum de ressources cognitives à l'association des phonèmes. Les bons et les mauvais lecteurs diffèrent en effet dans des tâches d'empan mémoriel, ces différences dérivant vraisemblablement de l'inefficacité du recodage phonétique en mémoire de travail.

Un moyen de compenser les difficultés du traitement phonologique est d'utiliser un ordinateur capable de prononcer les mots qui posent problème, ceci à la demande de l'utilisateur. Quelles sont les attentes que l'on peut formuler relativement à l'utilisation d'un tel système ?

Un système de rétroaction verbale peut en premier lieu être efficace pour permettre à l'enfant de corriger ses erreurs ou pour vérifier l'exactitude du décodage préalablement effectué. Si une erreur n'est pas identifiée, comme cela est souvent le cas en lecture silencieuse en classe, la mauvaise association graphème-phonème est renforcée. Cette association incorrecte peut par la suite interférer en mémoire avec d'autres relations rencontrées dans des mots orthographiquement similaires. Jorm et Share (1983) avancent que l'habileté à décoder correctement les mots fonctionne comme un mécanisme d'auto-enseignement, chaque identification correcte d'un mot constituant un apprentissage réussi qui mène à l'identification du mot à travers la voie directe. Il est en effet indiscutable qu'un enfant dont les traitements phonologiques sont efficaces développe un riche lexique orthographique : des opérations

réussies de décodage amènent à apparier des représentations phonologiques avec des représentations orthographiques ce qui permet un accès rapide au lexique mental.

Une rétroaction verbale peut également alléger la charge en mémoire de travail : les phonèmes peuvent être en effet correctement choisis, mais une capacité limitée de traitement peut empêcher que ceux-ci soient correctement assemblés. L'allégement de la charge en mémoire de travail produit de plus un meilleur stockage des mots : les connexions entre les sons et les lettres se trouvent renforcées ce qui facilite l'apprentissage de mots nouveaux par analogie avec des mots plus familiers (Ehri & Robbins, 1992).

La rétroaction verbale peut enfin développer certaines habiletés métacognitives. En effet, de façon à utiliser efficacement le système, le lecteur doit être conscient de la nécessité de demander ou non de l'aide; il doit donc être capable d'évaluer à tout moment le niveau de sa compréhension. Olofsson (1992, exp.1) étudie, chez des enfants de 81/2 ans, ce comportement de demande d'aide et l'auto-évaluation de leur compréhension. Les enfants lisent des textes soit sur papier soit sur écran d'ordinateur. Dans cette dernière condition, le lecteur peut pointer la souris sur un mot pour demander une rétroaction verbale : le mot est alors mis en surbrillance et prononcé rapidement. Après chaque texte, les enfants sont soumis à des questions de compréhension et, à la fin de la passation, à un questionnaire d'auto-évaluation. Les résultats montrent que les sujets demandent une rétroaction verbale pour des mots qui sont susceptibles de leur causer des difficultés de compréhension : les mots demandés sont généralement longs, irréguliers et de fréquence lexicale basse dans la langue. Il n'existe pas de différence au niveau de la compréhension des textes entre les deux groupes. Par contre, les sujets du groupe rétroaction verbale jugent leur compréhension à un niveau plus élevé que les sujets de l'autre groupe. Ces résultats sont en accord avec d'autres recherches étudiant, de façon plus générale, l'effet d'une rétroaction présentée sur ordinateur (*feedback* correctif). Par exemple, Zakay (1992) montre qu'une rétroaction sur ordinateur améliore le jugement qu'un individu peut avoir sur ses connaissances : la rétroaction permet de confirmer immédiatement les bonnes réponses ou de corriger les réponses erronées. Le fait que l'ordinateur produise la rétroaction permettrait à l'utilisateur, au contraire d'une interaction avec un adulte, de demander de l'aide sans retenue pour tester intentionnellement son savoir. Les ordinateurs sont en effet généralement perçus par les individus comme objectifs et moins embarrassants lors de l'émission d'une erreur. Les individus sont donc plus enclins à demander de l'aide à un ordinateur qu'à une personne (Kluger & Adler, 1993).

La rétroaction verbale peut donc faciliter les traitements phonologiques en améliorant le décodage et en fournissant ainsi à l'élève la possibilité de déléguer toute son attention à la compréhension. Une conséquence importante est que les élèves peuvent lire au-dessus de leur niveau en lecture. Les sujets testés par Olofsson (1992, exp.2) rapportent par exemple qu'ils lisent

davantage depuis leur entraînement et qu'ils sont moins effrayés pour lire des textes difficiles.

2. LA PRODUCTION DE LA PAROLE SUR ORDINATEUR

2.1 Données techniques

Il existe deux techniques pour restituer du son à partir de l'ordinateur : la numérisation et la synthèse du son.

La numérisation du son consiste à reproduire des sons enregistrés sur un support analogique ou par un microphone. Le signal électrique du son (produit par les vibrations de la membrane du microphone par exemple) est codé en une suite de valeurs numériques après avoir été découpé à des intervalles de temps réguliers (échantillonnage). Le son numérisé est alors stocké sur support informatique (disque dur, CD-ROM). La restitution du son nécessite que les opérations inverses soient effectuées : les données numériques sont converties en variations d'un signal électrique transmis, après amplification, à un haut parleur. Cette technique restitue la parole à la perfection; elle est de plus parfaitement compréhensible. Cependant, elle comporte un inconvénient : la très grande taille des fichiers de stockage doit être compensée par une vitesse rapide d'accès au disque dur pour que le délai entre la demande de rétroaction verbale et la production du son ne soit pas trop important. McConkie et Zola (1985) (cité in Farmer, Klein & Bryson, 1992) montrent à cet effet que, une fois que les mots ont été sélectionnés pour une rétroaction, le temps de recherche du mot peut varier de 500 ms à 10 secondes selon les mots. Selon les auteurs, cet intervalle peut décourager les lecteurs et réduire ainsi la portée du système.

La plupart des équipes de recherche utilisent un système de synthèse de textes. Le signal vocal est créé à partir de chaînes de caractères ASCII à l'aide d'un dictionnaire d'éléments acoustiques minimaux enregistrés et numérisés au préalable. Ce dictionnaire contient l'information phonologique et morphologique des items lexicaux ainsi que des données syntaxiques, sémantiques et pragmatiques du discours. Une des meilleures méthodes de transcription est une synthèse par diphones [7]. Cette méthode permet de traiter le difficile problème de coarticulation entre phonèmes. Un diphone est une réalisation acoustique qui s'étend de la zone stable d'un phonème à la zone stable du phonème suivant. On en compte 1200 en français d'une durée moyenne de 100 ms. L'enchaînement des diphones est produit en temps réel grâce à des

7 Cette méthode a été développée en France par le Centre National d'Etudes des Télécommunications (CNET). Deux systèmes importants ont été commercialisés sous licence CNET en France : le système Ferma F5000 et le système Televox (cf. Cohen, 1992, pour des exemples d'utilisation).

logiciels de transcription fonctionnant en arrière plan : la production du son est donc quasi instantanée. Cette méthode permet enfin de produire certaines caractéristiques suprasegmentales du texte lu : rythme, intonation, pause... Par exemple, le Ferma F5000 prononce différemment une phrase en fonction de la ponctuation du texte (*«non!»*, *«non ?»*), résout certaines ambiguïtés de prononciation (*«les poules du couvent couvent»*) grâce à des règles d'analyse syntaxique et connaît la prononciation de mots irréguliers (*«oignon»*, *«femme»*).

Le système largement utilisé dans les travaux rapportés ci-après est *DECtalk* développé à partir des recherches du *Massachusetts Institute of Technology* par *Digital Equipment Corporation*[8]. *DECtalk* est utilisé aux USA par l'équipe de Wise, Olson et collaborateurs et au Canada par Leong. Dans les pays scandinaves, une variante muti-langue de *DECtalk -InfoVox-* est utilisée notamment par Olofsson.

2.2 La compréhension de la parole de synthèse[9]

La perception de la parole de synthèse diffère de celle de la parole humaine par son insuffisance de prosodie (intonation, hauteur mélodique, durée de la parole) qui est la conséquence des routines stéréotypées d'articulation utilisées pour segmenter le message. De plus, dans le langage naturel, les distinctions entre les phonèmes sont spécifiées relativement à certaines données acoustiques fournies par les organes de la parole (par exemple, le timbre d'une voyelle est donnée relativement à la position de la langue et des lèvres et par le degré d'ouverture de la bouche). Ces données acoustiques sont implémentées en nombre limité dans la parole de synthèse ce qui produit une perte d'information non négligeable. Une conséquence importante est que les auditeurs compenseraient les défauts du signal en s'appuyant massivement sur le contexte pour accélérer la reconnaissance des mots, se trouvant ainsi dans la même situation qu'un lecteur ayant des difficultés de décodage (cf. supra Stanovich, 1980). Il convient donc de rechercher si, relativement à la parole normale, les traitements cognitifs de perception et de compréhension de la parole de synthèse sont affectés et modifiés par les caractéristiques du signal.

De nombreux résultats montrent effectivement que ces caractéristiques suscitent un coût cognitif supplémentaire; coût qui n'est apparent que si on utilise des mesures appropriées pour le mesurer. Une première mesure est l'exactitude des identifications ou l'intelligibilité du système[10]. Bien que l'intelligibilité du son restitué soit inférieure à celle de la parole enregistrée, il est

8 Pour une description détaillée du système, voir Bruckert, Minow & Tetschener (1983) et Klatt (1987). Voir également Venezky et Suraj (1993) qui présentent de nouveaux algorithmes pour présenter des mots en syllabes et les faire prononcer une par une par *DECtalk*.

9 Voir Duffy et Pisoni (1992) pour une synthèse détaillée de cette problématique.

10 Cette mesure est généralement appréhendée en demandant à des sujets de reconnaître un mot parmi plusieurs proposés après qu'ils aient été prononcés par le synthétiseur.

avéré que la plupart des systèmes de synthèse du son permettent une identification acceptable des mots de la langue. On peut noter cependant que ce paramètre varie considérablement en fonction du système utilisé (de 3% à 36% d'erreurs chez Logan, Greene & Pisoni (1989) qui comparent 10 systèmes en langue anglaise auprès d'adultes). *DECtalk*, dans sa version *Perfect Paul*, est le système le plus intelligible à l'heure actuelle. Olson, Foltz et Wise (1986) ont testé *DECtalk* auprès d'enfants et d'adultes. Un groupe de mauvais lecteurs répètent correctement 95% des 120 mots prononcés hors contexte par *DECtalk* et un groupe de sujets adultes répètent correctement 96% (respectivement, leur pourcentage est de 98% et 96% pour des mots préalablement enregistrés). Selon les auteurs, l'intelligibilité du système est suffisante pour permettre une rétroaction verbale de mots en contexte [11].

Le principal inconvénient de la parole de synthèse provient du fait qu'elle augmente le temps d'identification des mots. Pisoni (1981) utilise une tâche de décision lexicale et compare les temps de réactions obtenus pour des mots et des non-mots présentés auditivement à partir soit de la parole enregistrée soit de la parle synthétisée. Les résultats font apparaître que le temps pour décider du statut lexical du mot augmente avec de la parole synthétisée (de 153 ms pour les mots et de 133 ms pour les non-mots). L'absence d'interaction entre les deux facteurs laisse penser que le surcoût temporel est imputable à l'identification des phonèmes plus qu'à l'identification d'une entrée lexicale dans le lexique mental. Une étude de Manous et Pisoni (1984) indique de plus que les auditeurs doivent entendre plus de phonèmes pour identifier correctement un mot prononcé avec un système de synthèse. Les sujets écoutent successivement les 50, 100, 150 premières millisecondes d'un mot jusqu'à ce qu'il soit complètement présenté. Après chaque présentation, les sujets doivent noter le mot qu'ils pensent avoir reconnu. Les résultats indiquent que, quand le mot a été synthétisé (ici par *DECtalk*), les sujets ont besoin d'une durée de présentation de 417 ms en moyenne alors que cette durée minimale est de 361 ms quand le mot a été enregistré.

Il apparaît enfin que le traitement de la parole de synthèse, même si elle est clairement intelligible, nécessite plus de ressources cognitives en mémoire de travail que la parole humaine. Smither (1993) demande à ses sujets d'écrire plusieurs suites de huit chiffres après qu'elles aient été présentées auditivement soit par l'intermédiaire d'un système de synthèse soit par un magnétophone. L'intelligibilité du système a été précédemment testée auprès d'autres sujets qui ont reporté avec exactitude près de 99% des chiffres prononcés par le système. Les résultats font apparaître que les sujets réussissent mieux quand la parole est enregistrée que lorsqu'elle est synthétisée.

Il convient cependant de relativiser ces inconvénients car leurs effets sont atténués si l'utilisateur a suffisamment de pratique avec le système (Greenspan & al., 1988; Schwab, Nusbaum & Pisoni, 1985). Les résultats de

---

11 D'autres tests auprès d'adultes donnent sensiblement les mêmes résultats (Greene, Logan & Pisoni, 1986; Greenspan, Nusbaum, & Pisoni, 1988).

ces études font apparaître qu'avec la pratique (10 heures en moyenne), les utilisateurs améliorent leur habileté à reconnaître des mots synthétisés et qu'ils ne le font pas en apprenant à utiliser le contexte mais en devenant plus efficaces dans leur analyse du signal émis.

L'avantage majeur de la synthèse vocale reste sa souplesse d'utilisation : l'utilisateur d'un système de synthèse de textes n'a qu'à saisir, à l'aide d'un éditeur, la chaîne de caractères qu'il veut faire prononcer à l'ordinateur. Cependant, la voix restituée est artificielle. Comme le note Plaisant (in Cohen, 1992), cette imperfection de la voix a l'avantage de faire apparaître aux enfants qu'il ne s'agit que d'une machine; les interactions sont ainsi facilitées car elles ne revêtent pas le caractère complexe des communications que l'enfant développe avec les adultes.

3. PARADIGMES UTILISÉS DANS LES RECHERCHES SUR LA RÉTROACTION VERBALE

Les recherches qui testent l'efficacité des systèmes de rétroaction verbale sur l'amélioration de certaines habiletés en lecture sont apparues relativement récemment dans la littérature scientifique de psychologie cognitive (fin des années 1980). En conséquence, peu d'entre elles abordent la problématique de façon précise et certaines méthodologies utilisées laissent encore à désirer. Il a donc fallu faire un choix parmi elles, choix d'autant plus difficile à effectuer que le nombre de travaux est limité.

La plupart tentent de répondre à trois questions.

La première concerne l'étude du comportement des utilisateurs devant un dispositif de rétroaction verbale. Le comportement de demande d'aide est alors étudié en liaison avec certaines capacités métacognitives.

La seconde pose le problème de l'amélioration des habiletés testées. Pour étudier cette question, les expérimentateurs recherchent des *effets d'apprentissage* [12]. Pour cela, ils soumettent d'abord les utilisateurs à un *pré-test* dans lequel sont administrés des tests standardisés ou des épreuves construites par les expérimentateurs. Puis le sujet est soumis à une *phase d'utilisation* du système (la période étant très variable selon les recherches). En pratique, le sujet utilise un dispositif de pointage (souris le plus souvent mais également crayon optique ou contact avec l'écran) pour désigner ce qu'il veut faire prononcer par l'ordinateur. Celui-ci fournit alors le signal selon plusieurs formats (phonèmes, syllabes, mots, phrases). Dans certains cas, le sujet choisit lui-même le format qu'il désire entendre. Au fur et à mesure de l'émission du signal, les unités prononcées sont mises en surbrillance de façon à associer le traitement orthographique au traitement phonologique. Durant cette phase d'utilisation, plusieurs mesures comportementales sont recueillies : nombre

12 Les mots en italiques seront utilisés, sans explications, dans les descriptions des expériences.

de demandes d'aide, type de demande (quand le sujet a la possibilité de choisir la rétroaction verbale), latence des demandes, type de mots ou de segments de textes demandés. Enfin, la dernière phase est un *post-test* qui est destiné à évaluer les progrès accomplis depuis le pré-test. Les résultats sont également analysés en rapport avec les performances d'un groupe contrôle (ayant les mêmes caractéristiques que les sujets dits expérimentaux soumis à la rétroaction verbale) pour lequel aucune rétroaction verbale est donnée. Les effets d'apprentissage sont évalués en comparant les performances du pré-test et du post-test. Par exemple, les sujets peuvent être soumis durant le pré-test à une épreuve de décodage : ils lisent des mots à voix haute et l'expérimentateur note l'exactitude des prononciations et, dans certains cas, leur latence. Au post-test, les enfants sont confrontés aux mêmes mots ou à des mots aux caractéristiques similaires; les mesures comportementales sont alors comparées. Pour avoir une indication sur le *transfert d'apprentissage*, d'autres mots peuvent être présentés, généralement plus difficiles à décoder.

La dernière question concerne le type de segmentation à utiliser pour améliorer de façon optimale les habiletés considérées. Faut-il présenter le mot en entier ou le segmenter en syllabes ? Le type de segmentation optimal est-il fonction de la nature des déficiences en lecture des sujets ? La très importante recherche d'Olson et Wise (1992) aborde ce problème (cf. 4.1) : elle illustre le fait que la recherche dans ce domaine arrive enfin à maturité et qu'elle aborde, avec une méthodologie expérimentale sérieuse, des problématiques importantes qui ont des incidences sur la compréhension des mécanismes impliqués dans l'apprentissage de la lecture.

A ces trois questions, il convient d'ajouter celle qui concerne l'étude différentielle des populations testées : généralement, un groupe de sujets bons lecteurs est comparé à un ou plusieurs groupes de sujets mauvais lecteurs; des groupes de sujets avec des déficits sévères en lecture peuvent être également contrastés. En croisant diverses conditions expérimentales -possibilité ou non de rétroaction verbale, format de rétroaction proposé, population testée- les auteurs tentent de savoir s'il existe un type de segmentation optimal à proposer selon que l'utilisateur a ou non un déficit en lecture et selon la nature de ce déficit.

Certaines questions annexes ont déjà été abordées précédemment et nous n'y reviendrons pas (la compréhension de la parole de synthèse par exemple). D'autres, le comportement de demande d'aide et l'étude des capacités métacognitives requises à l'utilisation de systèmes de rétroaction verbale, seront encore développées ci-après.

La cohérence des recherches peut être appréhendée uniquement en fonction de la nature des habiletés cognitives en lecture que les auteurs cherchent à améliorer. C'est pourquoi nous avons choisi ce découpage -amélioration des capacités de décodage et de reconnaissance des mots; amélioration des capacités de compréhension- pour présenter les résultats de ces travaux.

4. AMÉLIORER LE DÉCODAGE ET LA RECONNAISSANCE DES MOTS

4.1 Le choix du format de la rétroaction verbale

L'amélioration du décodage et de la reconnaissance des mots pose le problème du format de la segmentation à utiliser pour donner une rétroaction verbale.

Pour van Daal et Reitsma (1990), le format de la rétroaction doit être en liaison avec les unités linguistiques utilisées lors de la lecture. Pour les mots plurisyllabiques, les auteurs suggèrent d'utiliser la syllabe, celle-ci étant l'unité de traitement la plus efficace en mémoire de travail et une unité compréhensible même pour des enfants dont les capacités métaphonologiques sont pauvres [13] (Baddeley, 1978; Fox & Routh, 1980; Rozin & Gleitman, 1977).

D'autres segmentations du mot sont possibles en particulier en groupes à centre vocalique (*GCV*, Spoehr & Smith, 1973; par exemple (mai/son; jar/din) et en fonction de la *BOSS* [14] (Taft, 1979), c'est à dire la structure syllabique orthographique de base qui préserve les unités morphémiques dans des mots plurimorphémiques (par exemple chant/er; coton/neux).

La syllabe peut être elle-même segmentée en attaque-rime dans le cas de mots monosyllabiques. La segmentation peut être également effectuée après la voyelle (par exemple pour *note* : /no/ + /te/). Cette dernière segmentation (ici appelée *segmentation post-voyelle*) semble produire de meilleurs apprentissages que celle en attaque-rime (Fayne & Bryant, 1981).

Quatre formats de rétroaction verbale sont généralement manipulés pour étudier leur efficacité sur le décodage et la reconnaissance des mots : la rétroaction peut concerner soit le mot entier, soit les syllabes du mot avec une segmentation généralement en *BOSS*, soit les unités subsyllabiques avec une segmentation en attaque-rime ou en post-voyelle, soit les phonèmes du mot.

Van Daal et Reitsma (1990) font l'hypothèse que, comparativement à une condition mot en entier, la segmentation en syllabes doit être préférée quand les problèmes de lecture se situent essentiellement dans le traitement grapho-phonologique. Ce type de segmentation peut stimuler le lecteur car il doit lui-même assembler les sons émis pour se représenter le patron sonore du mot entier. La condition mot entier peut, quant à elle, faciliter l'accès au lexique car le sens du mot est récupéré directement, améliorant en cela l'établissement des relations grapho-sémantiques. Dans leur étude, les sujets peuvent choisir entre deux types de rétroaction durant la phase d'utilisation [15] : soit le mot entier, soit des unités subsyllabiques (selon les mots, attaque-rime

13 Voir Gombert (1990), chapitre II, pour une revue de travaux portant sur le développement de la compétence à discriminer les sons langagiers.

14 En anglais pour *Basic Orthographic Syllabic structure*.

15 Les auteurs utilisent de la parole enregistrée.

ou post-voyelle). Les sujets sont des enfants de 10 ans ayant un retard en lecture de deux niveaux au moins. Les résultats montrent que, quel que soit le format choisi, les sujets demandent volontairement une rétroaction pour 46% des mots présentés. Cependant, le nombre de demandes dépend fortement du type de format : les demandes sont très élevées pour le mot entier (75%). En fait, ce comportement semble refléter une certaine paresse de la part des sujets qui leur est cependant très profitable. En effet, les latences des demandes sont plus élevées quand le choix porte sur la segmentation subsyllabique : tout se passe comme si l'enfant cherchait d'abord à décoder le mot puis, devant ses essais infructueux, demandait de l'aide. Cette hypothèse est vérifiée par l'étude des effets d'apprentissage. Un effet d'apprentissage est trouvé en ce qui concerne le décodage des mots, cet effet étant de même ampleur quel que soit le type de demande. Autrement dit, bien que la demande en subsyllabes soit beaucoup moins élevée que celle en mot entier, elle produit le même effet d'apprentissage indiquant en cela l'efficacité de ce type de rétroaction. De plus, il existe un transfert d'apprentissage plus important dans cette condition que dans la condition mot en entier. Cependant, les enfants ne sont pas très sélectifs dans leur demande : les mots qu'ils ont décodés sans erreurs durant le pré-test sont autant demandés que les mots pour lesquels ils ont échoué. Ce résultat pose le problème du comportement de demande d'aide, problème fondamental dans l'utilisation d'un système de rétroaction sur lequel nous reviendrons.

A notre connaissance, une seule recherche (Wise[16], 1988 citée in Olson, Wise & Rack, 1989) compare un format de rétroaction en *GCV* contre un format en *BOSS*. Le matériel est composé de mots plurisyllabiques et il est présenté à des enfants de niveau 2 en lecture. Ceux-ci, durant les trois phases de l'expérience, doivent prononcer à voix haute les mots; leur pourcentage de réponses correctes est ensuite analysé. Les résultats montrent un effet d'apprentissage plus important pour le format *BOSS*.

Dans une recherche très bien contrôlée au niveau méthodologique, Wise (1992) compare 4 formats de rétroaction : 1) le mot entier, 2) la syllabe (*BOSS*), 3) l'attaque-rime, 4) le phonème. Des sujets de 7 ans et de 8 ans sont divisés en trois groupes : niveau 1 d'apprentissage de la lecture avec performances normales à un test standardisé (1A - Normal), et niveau 1 et 2 avec de faibles performances (respectivement 1A - Bas et 2A - Bas). Le pré-test et le post-test sont constitués d'une épreuve de lecture à voix haute de 36 mots. Pendant la phase d'utilisation, l'enfant lit également le mot à voix haute après avoir demandé une rétroaction. Le pourcentage de réussite est analysé pour les trois phases.

16 C'est l'équipe de Wise et Olson, aux USA, qui a exploré en détail le problème de l'efficacité du type de segmentation. Sur la base des résultats d'un programme d'étude sur les dyslexies (*«Colorado remediation project»*, DeFries et al., 1991), ces auteurs ont été chargés de construire un outil susceptible de développer les habilités de reconnaissance des mots et de décodage phonologique chez des enfants dyslexiques.

La figure 1 montre l'effet d'apprentissage obtenu pour les trois groupes en fonction du format de rétroaction.

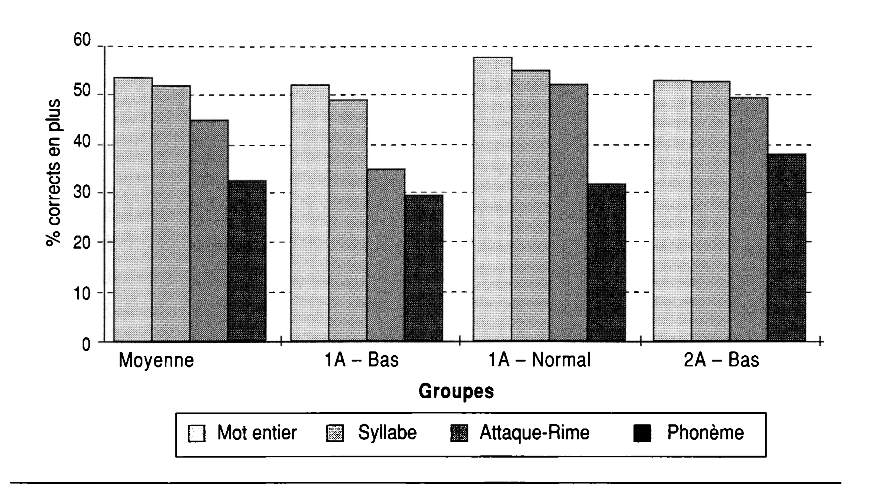


FIGURE 1

Différences post-test moins pré-test dans le pourcentage des mots correctement prononcés en fonction du type de rétroaction et du groupe [17]

On observe que l'effet d'apprentissage est faible lorsque le format concerne le phonème. Les sujets les plus jeunes ayant des difficultés de lecture (1A - Bas) ont plus de difficulté lorsque le format est segmenté en attaque-rime que lorsque le format concerne la syllabe ou le mot entier. Les autres sujets réussissent aussi facilement pour les formats segmentés (syllabe ou attaque-rime) que pour le format mot entier.

La charge en mémoire de travail explique la difficulté des enfants à associer plusieurs phonèmes pour former un mot et ceci d'autant plus que les mots sont longs. La figure 2 montre les résultats obtenus pour le groupe des enfants du groupe 1A - Bas en fonction de la longueur des mots.

Pour les mots monosyllabiques, le format attaque-rime n'est pas plus difficile que le format mot entier. Par contre, les enfants ont plus de difficulté avec ce format quand les mots sont plurisyllabiques, les effets d'apprentissage étant aussi faibles que ceux obtenus pour le format phonème. Quant au format syllabe, il produit des effets d'apprentissage sensiblement équivalents à ceux du format mot entier.

17 D'après Wise 1992, Fig.4 p.160

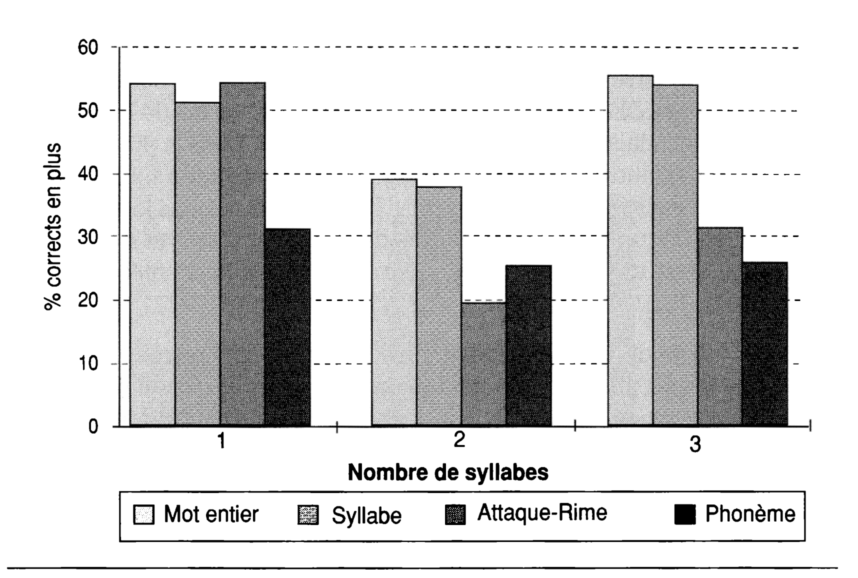


FIGURE 2

Groupe 1A - Bas : différences post-test moins pré-test dans le pourcentage des mots correctement prononcés en fonction du type de rétroaction et du nombre de syllabes [18].

Il ressort de ces résultats que 1) le format phonème est à proscrire [19]; 2) les formats en syllabe et en attaque-rime produisent des effets d'apprentissage semblables à ceux obtenu pour le format mot entier; 3) le format attaque-rime apparaît aussi difficile que le format phonème chez des lecteurs débutants ayant un faible niveau en lecture quand les mots sont longs.

Les recherches rapportées précédemment ont essentiellement contribué à sélectionner le format de rétroaction verbale à utiliser. Dans celles-ci, les sujets sont soumis à la rétroaction verbale durant une brève période de temps (une semaine au plus). Les effets à court terme obtenus permettent donc de déterminer quel est le format de rétroaction qui donne un apprentissage rapide et relativement stable. Des études à long terme sont cependant nécessaires pour évaluer si la rétroaction permet des transferts d'apprentissage. Dans ce type de recherches, les sujets sont soumis à la rétroaction pendant une période de plusieurs mois.

18 D'après Wise 1992, Fig.5 p.160.

19 L'expérience de Spaai, Reitsma et Ellermann (1991) conforte cette conclusion : ils observent chez des enfants allemands qui avaient 9 mois seulement d'apprentissage de la lecture que l'effet d'apprentissage pour le format phonème est significativement moins important que l'effet obtenu pour le format mot entier. De plus, cet effet ne diffère pas significativement de l'effet d'apprentissage obtenu pour le groupe sans rétroaction.

4.2 Les effets à long terme de la rétroaction verbale

Deux recherches, parmi les plus importantes, ont été menées par l'équipe de Wise et Olson pour étudier les effets à long terme de la rétroaction verbale sur le décodage et la reconnaissance des mots. Dans la première (Wise & al., 1989), un monitorat et des encouragements importants sont fournis par l'expérimentateur pour que l'enfant utilise la rétroaction. Dans la seconde (Olson & Wise, 1992), un plus grand éventail de sujets est utilisé et ceux-ci reçoivent un monitorat moins important. Les résultats des deux recherches sont comparés dans l'article de 1992 [20].

Des sujets de 10 ans en moyenne sont soumis à la rétroaction verbale. Ils lisent des textes pendant une demi-heure par jour pendant 6 mois. Les sujets sont sélectionnés à partir de tests standardisés évaluant leur niveau en lecture. Sur la base de ces résultats, les auteurs retiennent les sujets qui ont des déficits en lecture et les classent en deux groupes : un groupe de sujets avec un déficit léger en lecture et un groupe avec un déficit sévère.

En plus de ces tests, d'autres épreuves sont soumises aux sujets durant le pré-test et le post-test : une épreuve de reconnaissance de mots à temps limité (les mots apparaissent pendant 2 secondes à l'écran et l'enfant doit les prononcer correctement), une épreuve de décodage de non-mots; deux épreuves métaphonologiques (les sujets doivent composer des non-mots à partir de la première et dernière syllabe de deux mots et ôter un phonème dans un non-mot); et enfin un test d'intelligence.

Pendant la phase d'utilisation, les enfants choisissent eux-mêmes les textes qu'ils veulent lire (ceux-ci correspondent à leur niveau en lecture tel qu'il est évalué durant le pré-test). Après chaque texte, la compréhension est testée par des questions à choix multiples. Quatre conditions sont manipulées : les sujets reçoivent un format de rétroaction qui concerne soit 1) le mot entier; 2) la syllabe (*BOSS*); 3) l'attaque-rime. La dernière condition est la condition contrôle : les sujets sont soumis aux mêmes pré- et post-tests mais reçoivent normalement leur enseignement en classe.

En ce qui concerne le décodage de non-mots (cf. Figure 3), les résultats font apparaître que, relativement aux sujets du groupe contrôle, les sujets des groupes expérimentaux montrent un effet d'apprentissage 4 fois plus important. Le monitorat (phase 1) est bénéfique aux sujets notamment quand le format concerne l'attaque-rime et la syllabe. Si le monitorat est moins important, les gains obtenus dans les 3 formats sont équivalents. En moyenne, les sujets qui obtiennent les effets d'apprentissage les plus importants sont ceux qui ont reçu le plus d'attention individuelle et qui ont débuté leur apprentissage avec des scores élevés aux épreuves métaphonologiques.

---

20 On utilisera la notation d'Olson et Wise (1992) pour spécifier les recherches lors de la comparaison : la phase 1 concernera l'expérience de 1989 (monitorat important); la phase 2, celle de 1992 (monitorat moins important).

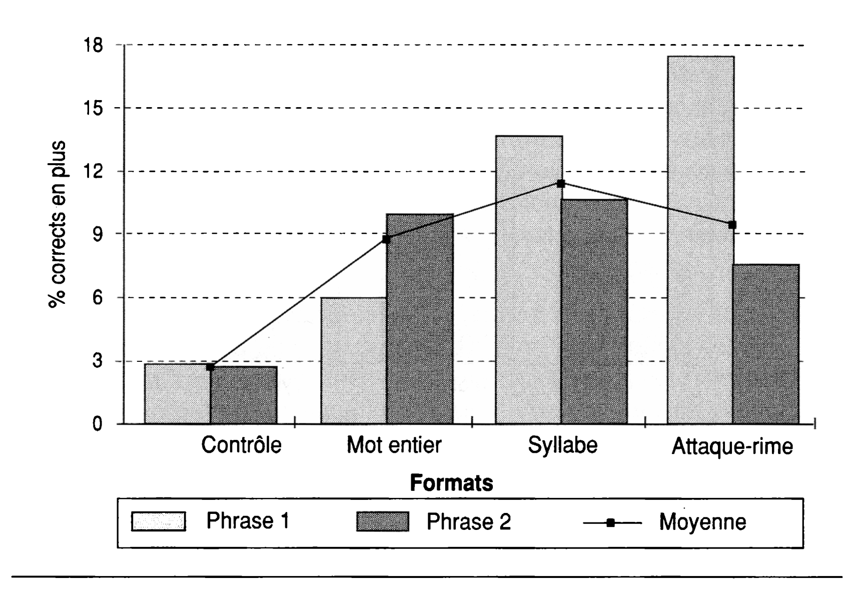


FIGURE 3

Différences post-test moins pré-test dans le pourcentage de non-mots correctement décodés en fonction du type de rétroaction et de l'importance du monitorat[21].

En ce qui concerne l'épreuve de reconnaissance de mots (cf. figure 4), l'effet d'apprentissage des groupes expérimentaux est deux fois plus important que celui du groupe contrôle. L'effet du monitorat s'avère ici non significatif et il n'existe pas d'interaction avec le format. Cependant, une différence significative est observée entre les 2 phases pour le format attaque-rime.

Sur les deux phases, le niveau en lecture des sujets du groupe contrôle augmente de 0.33 niveau alors que le gain est de 0.66 niveau pour les sujets expérimentaux. L'ensemble de ces résultats doit néanmoins être analysé en fonction du degré de sévérité des déficits en lecture des sujets testés.

Par exemple, en ce qui concerne le décodage de non-mots (cf. figure 5), c'est le format attaque-rime qui produit les meilleurs effets d'apprentissage pour les sujets avec un déficit léger. Pour les sujets avec un déficit sévère, le format syllabe (BOSS) est le meilleur. Pour ces sujets, l'effet d'apprentissage obtenu pour le format syllabe est 140% supérieur à celui obtenu pour le format attaque-rime et 61% supérieur à celui obtenu pour le format mot entier.

21 D'après Olson & Wise 1992, Tab.3 p.123.

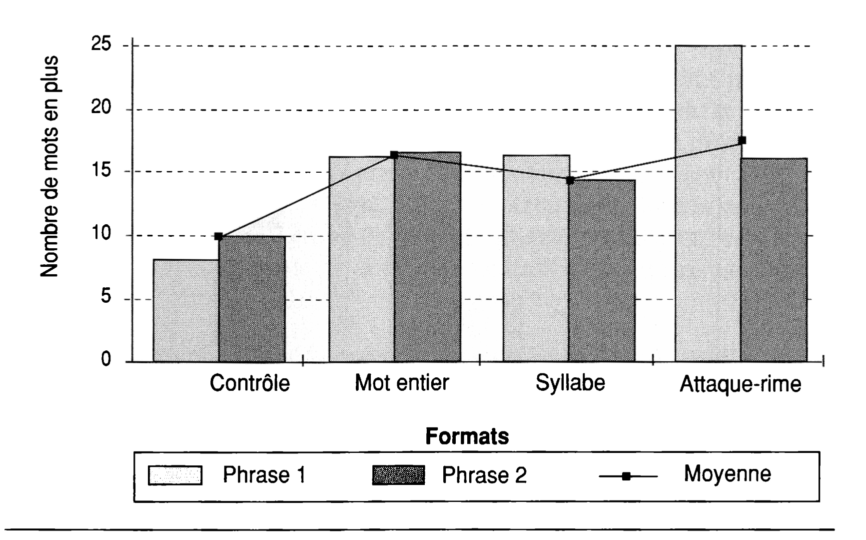


FIGURE 4

Différences post-test moins pré-test dans le nombre de mots correctement lus en fonction du type de rétroaction et de l'importance du monitorat[22].

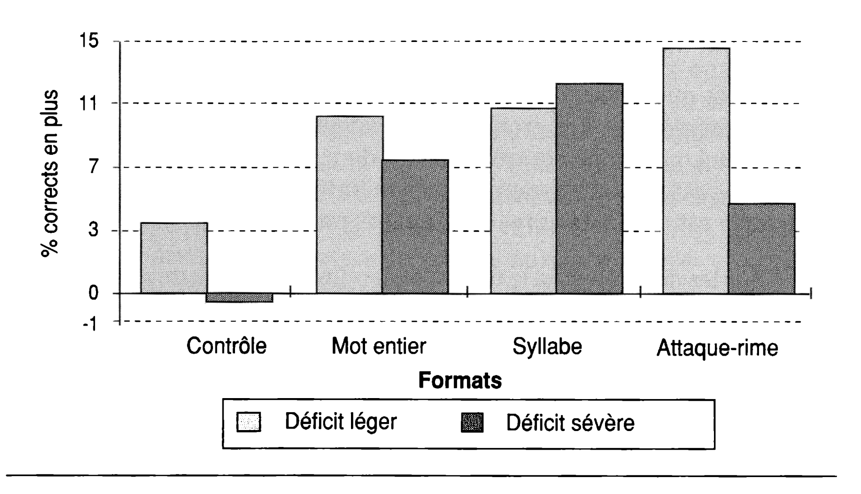


FIGURE 5

Différences post-test moins pré-test dans le pourcentage de non-mots correctement décodés en fonction du type de rétroaction et du degré de sévérité du déficit en lecture. [23]

22 D'après Olson & Wise 1992, Tab.4 p.125.

23 D'après Olson & Wise 1992, Tab.5 p.127.

En ce qui concerne les épreuves métaphonologiques, les résultats indiquent qu'il existe une relation forte entre le score obtenu à celles-ci lors du pré-test et l'importance des effets d'apprentissage sur le décodage et la reconnaissance des mots. Les sujets qui sont au-dessus de la moyenne à l'épreuve de construction de non-mots ont des effets d'apprentissage sur le décodage deux fois plus importants que les sujets en dessous de la moyenne. De plus, on observe chez les sujets qui réussissent le mieux l'épreuve de délétion de phonèmes, des effets d'apprentissage plus importants sur la reconnaissance des mots.

On observe donc des effets à long terme d'une rétroaction verbale sur le décodage et la reconnaissance de mots. De plus, les résultats de Wise et collaborateurs montrent l'importance du monitorat et des encouragements donnés par l'expérimentateur dans l'utilisation d'un tel outil de remédiation. L'effet du format de la rétroaction varie en fonction du déficit en lecture : le format attaque-rime est le plus avantageux pour des sujets avec un déficit léger mais c'est le plus handicapant pour des sujets avec un déficit sévère. Le format syllabe est le plus avantageux pour des sujets avec un déficit sévère et il est équivalent au format mot entier pour des sujets avec un déficit léger. Enfin, les résultats sont liés aux capacités métaphonologiques des sujets : l'importance des effets d'apprentissage est prédite par la réussite aux épreuves métaphonologiques.

4.3 Le comportement de demande d'aide[24]

On a vu que les résultats de Van Daal et Reitsma (1990) posaient le problème de la demande d'aide. Les auteurs suggèrent que les mauvais lecteurs ne sont pas capables de déterminer quand ils ont réellement besoin d'aide.

La recherche d'Olofsson (1992, exp.2) conforte ce propos. L'auteur étudie l'effet d'une rétroaction verbale en mot entier chez des enfants mauvais lecteurs de niveau 2 (8 ans) à 7 (13 ans). Les résultats indiquent des effets d'apprentissage plus importants pour le décodage de mots et de non-mots quand l'enfant se situe à un niveau de lecture élevé (niveaux 4 à 7). Une période plus longue d'apprentissage du système est nécessaire chez les jeunes enfants. Ceux-ci ont également besoin que l'expérimentateur (ici leur enseignant) les guide dans l'utilisation du système et dans leur demande d'aide.

Il apparaît ainsi que l'utilisation d'un système de rétroaction verbale s'avère limité chez des sujets jeunes qui n'ont pas encore de réelles capacités métacognitives pour gérer correctement leur demande d'aide et pour auto-évaluer leurs connaissances. Farmer et al. (1992) observent le même résultat

24 Voir van Daal et van der Leij (1992) et Wise et Olson (1992) qui analysent les comportement de demande d'aide dans des expérimentations avec des enfants ayant des difficultés d'écriture. Dans ces recherches, le système de rétroaction, après avoir donné une rétroaction verbale, donne une rétroaction de ce que l'enfant écrit à l'aide du clavier de l'ordinateur.

pour des sujets ayant des déficits sévères en lecture. Leurs sujets sont âgés de 16 ans en moyenne et ont un niveau en lecture de 11 ans avec un QI cependant normal. Ils lisent des textes de 140 mots avec un format de rétroaction portant sur le mot entier. Aucun effet d'apprentissage est observé : les demandes sont en effet très peu nombreuses (4 mots en moyenne par texte), les sujets devant être constamment sollicités durant la phase d'utilisation pour exploiter efficacement le système.

Van Daal et Reitsma (1993) montrent des résultats sensiblement différents de ceux de 1992. Les auteurs font l'hypothèse que leurs résultats pourraient être biaisés du fait du faible pourcentage de mots faciles à décoder dans les textes : le rapport (de 1 mot facile pour 4 mots difficiles) aurait produit l'impression que tous les mots étaient difficiles à décoder. Ils égalisent donc le nombre de mots faciles et difficiles à décoder dans leur matériel. Trois groupes de sujets sont utilisés. Un groupe de lecteurs avec un niveau normal en lecture et un groupe de lecteurs ayant un retard de 1.5 niveau par rapport au premier groupe lisent des listes de mots (450 au total sur 8 sessions) avec une rétroaction verbale à la demande dans le format mot entier. Un autre groupe de sujets, ayant un retard en lecture, lisent en condition rétroaction verbale obligatoire. Les résultats montrent que les mauvais lecteurs sont ici plus prompts à demander de l'aide pour les mots difficiles; il n'existe pas, à ce niveau, de différence significative avec les bons lecteurs. Ces derniers demandent de l'aide essentiellement au début des premières sessions de la phase d'utilisation alors que l'autre groupe reste dépendant de la rétroaction verbale jusqu'à la fin des sessions. Un effet d'apprentissage est observé au niveau de l'exactitude du décodage quel que soit le groupe. Au niveau de la vitesse de décodage, un effet d'apprentissage est également obtenu mais il diffère selon le type de rétroaction : les bons lecteurs bénéficient plus de la rétroaction lorsque celle-ci est à la demande alors qu'une rétroaction obligatoire est plus profitable aux mauvais lecteurs.

Il semble donc que la rétroaction verbale doit être manipulée en fonction du degré de sévérité des déficits en lecture des utilisateurs. Manifestement, les bons lecteurs utilisent celle-ci, volontairement, pour évaluer l'exactitude de leur décodage. Par contre, l'inefficacité du décodage des mauvais lecteurs les empêche de détecter leurs erreurs : dans ce cas, une rétroaction verbale obligatoire leur est plus profitable.

5. AMÉLIORER LA COMPRÉHENSION

La rétroaction verbale peut faciliter la compréhension d'un texte en permettant au lecteur de mieux mémoriser les unités de traitement linguistique (mots, syntagmes, propositions et phrases), de mieux intégrer l'information nouvelle à l'information ancienne et de mieux « purger » l'information ancienne de la mémoire de travail pour l'encodage propositionnel (Leong, 1992a). Les résultats obtenus ne sont pas homogènes et il convient de les discuter en fonction de la population testée et du format de la rétroaction.

Leong (1992a) cherche à améliorer la compréhension de textes auprès d'enfants de niveaux 6, 7 et 8 à l'aide d'un système d'aide interfacé avec *DECtalk*. A chaque niveau de lecture, les enfants sont divisés en deux groupes après la passation de tests de compréhension standardisés : ceux qui sont au-dessus de leur niveau et ceux se situant en dessous de leur niveau d'apprentissage de la lecture. L'étude (d'une durée de 18 semaines) comprend un pré-test de lecture de 12 textes de 200 mots sur ordinateur, une phase d'utilisation du système avec la lecture des mêmes textes en auto-présentation[25] avec ou sans rétroaction verbale et un post-test. Le pré-test comprend également une familiarisation à l'utilisation du système et une évaluation de la compréhension des textes par des questions ouvertes inférentielles. Chaque texte est segmenté en propositions et présenté sous cette forme en auto-présentation. Dans la condition rétroaction verbale, *DECtalk* prononce les propositions au rythme de 200 mots par minute. Le sens de certains mots présélectionnés (mis en surbrillance) peut être demandé par les sujets dans les deux conditions expérimentales. Les scores obtenus aux questions de compréhension sont analysés. Les résultats font apparaître qu'il existe une amélioration de la compréhension pour les deux conditions de présentation mais l'efficacité du système est observée pour seulement deux textes chez des sujets se situant au-dessus de leur niveau attendu de compréhension.

Selon Leong, plusieurs paramètres sont à considérer pour améliorer l'efficacité du système dans la compréhension de textes. D'abord l'organisation textuelle, tant au niveau de la surface (longueur des mots, fréquence lexicale, familiarité) qu'à un niveau plus profond (complexité inférentielle), doit être travaillée pour rendre le matériel compréhensible. Ensuite, la présentation des textes sur ordinateur doit être faite sous un format qui facilite la prise d'information. Enfin, le niveau d'expertise des lecteurs doit être spécifié au niveau de leur savoir général, de leur savoir textuel et de leurs stratégies d'apprentissage.

Elkind, Cohen et Murray (1993) montrent des résultats plus tangibles que ceux de l'expérience précédente. Les auteurs étudient, auprès d'une population de 28 enfants dyslexiques, si un système de rétroaction verbale améliore leur habileté à comprendre des textes. Les sujets sont issus d'une école privée où ils sont soumis à un programme de remédiation. Pour 70% d'entre eux, leur niveau en lecture est inférieur d'au moins un niveau à celui attendu. Le système utilisé est un véritable système d'aide à la lecture : un mot préalablement sélectionné peut être prononcé en entier, syllabe par syllabe ou lettre par lettre; le sujet peut également faire prononcer une phrase entière, demander le sens d'un mot et augmenter ou diminuer la vitesse de prononciation. Les sujets sont divisés en deux groupes : l'un bénéficie du système pour lire, l'autre non. Les sujets subissent un pré-test en début d'année scolaire et un post-test un semestre plus tard (les tests sont des tests standardisés de com-

25 Le sujet appuie sur la barre espace de l'ordinateur pour faire apparaître les segments qui s'affichent au fur et à mesure sur l'écran.

préhension). Les résultats font apparaître que les sujets qui ont bénéficié du système d'aide (20 à 25 heures d'utilisation) augmentent leur niveau de compréhension en lecture de 1.2 niveau en moyenne. 70% d'entre eux montrent un gain d'au moins un niveau, 51% de ceux-ci gagnant de 2 à 5 niveaux. Tous les sujets ne bénéficient pas de l'utilisation du système : 14% ne montrent pas d'évolution et 14% réussissent moins bien après l'utilisation. Relativement à ce dernier résultat, les auteurs font observer qu'il convient d'être sélectif dans l'utilisation d'un tel système et qu'il est souhaitable de déterminer, par un diagnostic préalable des difficultés, les individus qui pourront bénéficier d'un système d'aide.

Conclusion

Bien que le nombre de recherches dans le domaine de la rétroaction verbale soit encore relativement restreint, il est certain que des résultats encourageants émergent de celles-ci. D'après ce que rapportent les auteurs des travaux, les enfants se réjouissent d'utiliser un tel système; de même, les enseignants et les parents reconnaissent que les enfants progressent et qu'ils s'intéressent plus à la lecture (Olson & Wise, 1992).

La rétroaction verbale a des effets à long terme sur l'efficacité du décodage, de la reconnaissance des mots et de la compréhension. Les résultats les plus intéressants concernent le format de la rétroaction à utiliser. D'une façon générale, une segmentation en unités subsyllabiques (en attaque-rime notamment) apparaît très profitable pour améliorer le décodage et la reconnaissance des mots. Ce type de format est particulièrement adapté à des enfants qui débutent l'apprentissage de la lecture et qui n'ont pas de déficits notamment au niveau des capacités métaphonologiques. Il permet lui-même que ces capacités métaphonologiques se développent facilitant en cela l'apprentissage de la lecture. La rétroaction verbale agit manifestement en allégeant la charge cognitive et, en conséquence, permet la constitution de relations fortes entre les représentations mentales des codes phonologiques et orthographiques. Il apparaît en effet que, lorsque la surcharge cognitive est trop importante dans une rétroaction, comme ceci est le cas pour une segmentation en unités phonémiques, le bénéfice apporté par celle-ci est nul. Pour les enfants plus âgés, un format mot entier est également profitable. Manifestement ce type de format est adapté à une période d'apprentissage pour laquelle les enfants reconnaissent de nouveaux mots par analogie avec des mots anciens.

Les résultats obtenus avec des enfants présentant des déficits en lecture sont également intéressants; le format de la rétroaction interagissant avec le degré de sévérité des déficits. Pour ce type de lecteurs, une segmentation en syllabes (*BOSS*) ou une présentation du mot entier apparaît profitable. Cependant, l'utilisation efficace du système pose problème. Une rétroaction obligatoire s'avère plus efficace pour eux. De plus, ces lecteurs

présentant des déficits au niveau des capacités métaphonologiques, il apparaît nécessaire de mettre en place des phases intensives de monitorat.

Plusieurs recherches ont en effet montré la nécessité d'incorporer, dans les phases initiales d'utilisation de systèmes à rétroaction verbale, des guidances métacognitives qui aident le lecteur à internaliser son savoir. Dans des systèmes d'aide à la lecture plus sophistiqués, des exercices aident le lecteur à être plus conscient de sa lecture et à produire plus d'efforts dans ses apprentissages. L'équipe de Scardamalia et Bereiter ont par exemple mis au point un système (*CSILE* pour *Computer-Suported Intentional Learning Environments;* cf. Scardamalia & al., 1989) qui peut être utilisé quel que soit le niveau de lecture. *CSILE* est un tuteur intelligent qui fonctionne à partir d'une base de données construite, en temps réel, à partir des actions émises par les utilisateurs. Ceux-ci participent donc activement à la construction du savoir implémenté en interagissant entre eux. Le système peut être utilisé indifféremment par des élèves, des enseignants ou des experts du domaine.

Il demeure que les recherches futures devront, d'une part, répliquer les résultats obtenus et, d'autre part, déterminer très exactement le type de système (rétroaction à la demande *versus* obligatoire) et le format de la rétroaction à utiliser selon le niveau en lecture des utilisateurs et la nature de leurs déficits.

L'ensemble de ces résultats est par ailleurs parfaitement en accord avec ce que l'on sait déjà de l'acquisition de la lecture : les capacités métaphonologiques et la maîtrise du décodage sont des aspects critiques dans l'acquisition du code alphabétique. On a vu que leurs pratiques étaient largement développées dans les systèmes de rétroaction présentés et qu'elles pouvaient être améliorées. En cela, ce courant de recherches apporte des évidences supplémentaires sur l'importance de l'apprentissage systématique du code alphabétique dans l'enseignement de la lecture.

La synthèse vocale, comme nouveau moyen d'appropriation de l'écrit, ouvre donc des perspectives nouvelles dans l'enseignement de la lecture car elle permet à l'enfant de prendre en charge l'acquisition de ses connaissances. Il ne fait aucun doute que ces technologies nouvelles sont en passe de modifier profondément l'organisation du système scolaire. Il reste cependant aux législateurs de préparer cette mutation en fournissant à celui-ci les moyens appropriés et en agissant au niveau de la formation des maîtres.



LA MÉTACOGNITION



Évaluation formative et psychologie cognitive : mouvances et tendances

Gérard Scallon

Introduction

Il y a de ces disciplines, de ces sujets d'étude, qui présentent toutes les apparences d'un corpus de connaissances relativement stables. Par exemple, et sans vouloir offusquer personne ni diminuer l'effort de certains collègues, les enseignements relatifs aux caractéristiques et aux applications de la courbe normale des probabilités n'ont pas beaucoup changé depuis ces dernières années; ni d'ailleurs l'enseignement des algorithmes de résolution d'équations du premier degré à deux inconnues ! On ne peut en dire autant de la génétique, de la technologie informatique et de l'évaluation formative des apprentissages !

Au cours des deux dernières décennies, l'évaluation formative a fait l'objet de plusieurs écrits et, malgré tout, elle s'inscrit encore aujourd'hui dans le sillon d'une réflexion intense qui n'a jamais cessé de dominer la théorie et la pratique. Après les clarifications qui s'imposaient quant à sa nature et aux aspects multiples de sa fonction de régulation, on a peut-être eu l'impression de pouvoir se reposer un peu en laissant la poussière retomber. Mais «pierre qui roule n'amasse pas mousse» et la pratique de l'évaluation formative, là où on a réussi à l'implanter avec de nombreux procédés de séduction, est toujours une entreprise en chantier.

L'avènement des sciences dites cognitives et des pédagogies qui s'en inspirent vient raviver un domaine encore en ébullition. L'auteur de ce texte s'intéresse à la méthodologie de l'évaluation formative depuis plusieurs années et, au moment d'entrevoir les perspectives qui s'offrent à ce type d'évaluation, il a cru bon de prendre un certain recul afin de mieux entrevoir l'horizon. Les mouvances et tendances qui s'annoncent indiquent qu'il faudra bien tôt ou tard se mettre à pied d'œuvre car les impératifs de certaines pédagogies sont déjà en avance sur les pratiques d'évaluation qui ont été mises en œuvre.

1. L'HÉRITAGE DES PIONNIERS

Les premiers exemples qui nous ont été donnés comme source d'inspiration pour la méthodologie de l'évaluation formative s'inscrivent dans une pédagogie par objectifs teintée d'une conception de l'éducation que l'on a appelée «pédagogie du succès» (ou la pédagogie de la maîtrise qui est une traduction littérale du «mastery learning» des écrits anglo-saxons). Bloom, Hastings et Madaus (1971 et 1981) ont sans doute donné le coup d'envoi en proposant un exemple de questionnaire écrit suivi d'un corrigé constitué de renvois divers aux notes de cours ou au manuel de base. L'utilisation de ce corrigé par les élèves devait tenir lieu de correctif pédagogique ce qui confère un caractère formatif à l'instrument utilisé (Bloom et coll., 1971, p. 129 et suiv.).

Il serait difficile de rendre compte des approches méthodologiques qui se sont succédées par la suite. Les dispositifs d'enseignement qui ont servi

à rendre opérationnelle la pédagogie du succès, que ce soit pour accommoder l'enseignement collectif (p. ex. Block et Anderson, 1975) ou l'enseignement individualisé (p. ex., Keller, 1968), se sont prêtés à une méthodologie de l'évaluation formative qui présente des caractères constants : un contenu découpé en modules, une progression balisée en objectifs intermédiaires et en objectifs terminaux et des tests dits formatifs (de type papier crayon, pour la plupart). La mesure à interprétation critérielle, vouée plus ou moins explicitement à la certification des compétences en évaluation sommative (ce qui n'est pas toujours évident dans les écrits d'origine), a vite été récupérée par les praticiens de la pédagogie du succès pour devenir l'élément clé de la méthodologie de l'évaluation formative. La séquence *«apprentissage, évaluation formative, enseignement correctif»* est devenue le cadre conceptuel le plus répandu pour alimenter le discours sur l'évaluation formative que ce soit pour rédiger des guides docimologiques ou pour énoncer des politiques d'évaluation. Selon ce modèle de démarche, les trois éléments se présentent comme des étapes distinctes et, du point de vue de la formation des personnels de l'éducation, l'évaluation formative est en soi un objet d'étude qui se distingue de la didactique et de la pédagogie, ce qui n'est certes pas forcément la seule façon de voir comme en témoignent des écrits récents (p. ex., Bain, 1988; Allal, 1979 et 1988; Scallon, 1992). La notion de «remédiation» n'est cependant pas une notion abusée et peut se présenter comme un moyen de régulation approprié selon certaines théories d'apprentissage comme en témoigne un ouvrage récent de Fisher (1993, pp. 22 et suivantes). Dans la suite de ce texte, la remédiation comme telle n'est pas remise en cause. Il sera plutôt question de réexaminer le caractère disjoint des trois phases de la séquence «apprentissage, évaluation formative, enseignement correctif» tel que suggéré dans les écrits traditionnels sur l'évaluation.

2. LORSQUE LA NÉCESSITÉ EST MÈRE DE L'INVENTION

Qu'arrive-t-il si la situation pédagogique n'offre pas aux élèves un itinéraire balisé en objectifs ou en tranches de contenu ? Ou encore, qu'arrive-t-il lorsque le parcours ne s'échelonne plus depuis les objectifs les moins élevés d'une hiérarchie jusqu'aux objectifs terminaux ? Et c'est pourtant le genre d'interrogation qu'il faut soulever au regard des pédagogies dites de situation ou de projets ou encore dans des contextes d'apprentissage par problèmes. Les programmes révisés de français langue maternelle pour les écoles primaires du Québec (Ministère de l'Éducation du Québec, 1979), des approches nouvelles de la didactique de la lecture au primaire (Tremblay et Demers, 1990) et enfin l'apprentissage par problèmes que plusieurs facultés de médecine, dont celles du Québec, ont adopté et sont en voie de mettre en place (voir Berkson, 1990) en sont des exemples bien concrets. On comprendra que, dans ces divers cas, le point d'entrée de toute démarche d'apprentissage est une tâche complexe à réaliser ou un problème à résoudre, c'est-à-dire une tâche correspondant à un élément beaucoup plus près de l'objectif terminal

que d'un objectif de départ dans une longue séquence d'apprentissages intermédiaires. La séquence « apprentissage / évaluation formative / enseignement correctif » n'est plus appropriée pour répondre aux interrogations soulevées dans le cadre de pédagogies dites de situation.

D'abord, l'apprentissage n'est pas soumis aux règles de la pédagogie de maîtrise qui indiquent que tout élément doit être maîtrisé avant d'en aborder un autre. Dans cet ordre d'idées il est difficile de concevoir un enseignement correctif qui a pour visée principale le contrôle rigoureux d'une progression. Que dire alors de l'évaluation formative proprement dite ? L'incertitude n'est pas aussi évidente qu'on l'aurait cru car il est toujours possible d'observer, de mesurer ou encore d'apprécier une performance. L'imbroglio se présente avec netteté lorsqu'il s'agit de concevoir ou de visualiser la mise en scène. Qui évalue en pédagogie de situation ? La réponse est toute trouvée en imaginant un épisode d'objectivation de sa pratique de communication par l'élève dans le cadre des programmes renouvelés de français au Québec. La réponse vient naturellement aussi lorsque, dans une situation d'apprentissage par problèmes comme en médecine, les étudiants doivent se saisir d'un cas clinique et établir leurs besoins d'information et leurs objectifs d'apprentissage avant d'aller plus loin. La participation, voire l'engagement personnel des élèves prend de l'ampleur et vient agrandir l'écart qui s'est creusé entre un modèle d'évaluation formative de départ et une pratique d'évaluation qui doit répondre à des impératifs d'un tout autre ordre que ceux de la pédagogie de maîtrise. Il est important de constater que la participation des élèves à leur propre évaluation formative ne tient pas à des exigences d'ordre purement pratique telles l'allégement du fardeau que présente souvent l'évaluation dans un groupe d'élèves au regard de performances complexes, par exemple. Cette particularité tient beaucoup plus à un changement majeur au niveau de l'objet même de l'évaluation formative, un matériau nouveau que l'on trouve dans le champ des sciences cognitives. J'aurai certes l'occasion d'y revenir avec plus de détails.

En bref, le discours des dernières années sur l'évaluation formative et les préoccupations des praticiens de l'éducation ont permis de révéler que le modèle de départ à trois phases, soit l' "apprentissage-évaluation-correctif", n'était pas facilement généralisable à des contextes éducatifs autres que ceux qui ont servi d'assise à la pédagogie du succès.

3. VERS UNE LIBÉRATION DE L'ÉVALUATION FORMATIVE

La pédagogie du succès a exercé une telle séduction que l'on a presque cru pendant un certain temps qu'il y avait un lien inséparable entre ce type de pédagogie et l'évaluation formative des apprentissages. Pour ce qui est de briser ce lien, Allal (1978) a fait œuvre de pionnière en quelque sorte en posant la relation qui pouvait exister entre la méthodologie de l'évaluation

formative et les conceptions psychologiques ou psychopédagogiques de l'apprentissage. La séquence de départ n'apparaît déjà plus comme étant la seule façon d'intégrer l'évaluation formative aux pratiques pédagogiques puisque des démarches non instrumentées, de nature interactive pour ce qui est des procédés de régulation, peuvent devenir le lot de plusieurs contextes pédagogiques. Allal a donné le nom de régulation interactive à cette démarche de guidage continu de la situation d'enseignement et d'apprentissage pour mieux l'ajuster aux caractéristiques personnelles des élèves (1979, pp. 135 et 138).

Les principes de nature docimologique ont formé également une autre enclave pour l'évaluation formative. Scallon (1992) a tenté de démontrer que, pendant un certain temps, la méthodologie de l'évaluation formative a fait l'objet d'un discours aux préoccupations de nature purement docimologique comme la recherche du nombre minimal de questions à poser dans un exercice et la détermination du seuil idéal de réussite et ce, pour départager avec le plus de certitude possible les «élèves qui maîtrisent» de ceux «qui ne maîtrisent pas» ! D'ailleurs, très peu de procédés de détermination de cet indicateur de maîtrise, là où il s'applique, posent comme critère l'efficacité du correctif prescrit au regard des élèves qui y ont été assignés. Mais il y a bien plus ! Il est habituellement reconnu que lors d'un examen, d'un contrôle ou d'un exercice de vérification, le portrait de l'élève ne devrait pas être modifié ou altéré. Le principe est solidement inscrit dans les moeurs d'évaluation et provient sans aucun doute de préoccupations qui doivent caractériser toute opération de mesure. Par exemple, en calorimétrie on doit éviter que la chaleur d'un thermomètre altère la température que l'on veut mesurer pour un liquide. Voilà pour la préoccupation d'ordre docimologique lorsque le portrait à préserver est une compétence, une habileté ou un répertoire de connaissances.

La notion de rétroaction (information en retour ou «feed-back») peut se substituer à cette préoccupation. Les démarches de vérification qui proposent aux élèves un corrigé ou un solutionnaire pour vérifier leur performance sont de cet ordre. Les instruments conçus pour être utilisés par les élèves mêmes de manière à ce que ceux-ci soient placés aux premières loges de la rétroaction s'inscrivent nettement dans cette perspective. Les tâches de «surlignement» que les élèves doivent effectuer dans certains exercices de français écrit avant de fournir une réponse-produit (comme accorder des verbes avec leurs sujets, par exemple) sont également du même ordre (voir Tremblay et Demers, 1988 et 1990). On comprendra alors que la rétroaction, intégrée à l'ensemble de tâches ou de problèmes que le sujet doit réaliser ou résoudre, peut altérer au mieux le portrait de l'élève pendant le déroulement même de la démarche d'observation ou de prise d'information. Cette façon de voir contraste avec la perspective docimologique et présente plutôt des préoccupations de nature didactique ou pédagogique (Scallon, 1992). Et, pour ce qui est de soumettre l'évaluation formative à des contrôles de qualité, les recherches menées dans le cadre des diverses applications du feed-back ont permis d'asseoir cette pratique d'évaluation sur des bases beaucoup plus rationnelles qu'empiriques. On peut se permettre d'ajouter que lorsque le feed-

back se pose comme destination des actes d'évaluation formative, ceux-ci se trouvent libérés en quelque sorte du joug de l'enseignement correctif comme en témoignent certains écrits sur le sujet (p. ex., Yeany et Miller, 1983; Provencher, 1985).

Sans aucun doute, les conceptions cognitivistes de l'apprentissage ont posé, posent et poseront vraisemblablement des défis majeurs à la pratique de l'évaluation formative. Tout semble jusqu'ici s'être joué autour des scénarios ou des mises en scène pour ce qui est du rôle des acteurs engagés dans une situation d'enseignement et d'apprentissage. Nul doute que l'élève y joue un rôle primordial, ce qui ne peut être remis en question. Qu'en est-il de la structure de ce qui doit être observé ? Par exemple, le produit d'une démarche ou le processus même lié à cette démarche ? Et si le «quoi» de l'évaluation formative est relativement facile à déterminer, le «comment» l'est beaucoup moins. Il faut admettre que le discours cognitiviste, en dépit de la forte séduction qu'il a exercé et exerce encore dans les milieux éducatifs, est demeuré assez imperméable aux impératifs de l'évaluation formative, laquelle doit avoir pour point de départ un processus d'observation. Les définitions opérationnelles ne sont pas légion et il ne faut jamais perdre de vue que toute observation, voire toute mesure (avant d'arriver à l'évaluation) doit s'alimenter à des indicateurs de surface. Par exemple, la ponctualité n'est pas visible comme telle mais les conduites qu'elle entraîne chez les individus le sont. Comme autre exemple on ne peut saisir directement l'état d'anxiété d'un individu. Tout au plus, peut-on dénombrer des comportements indicateurs d'anxiété. Cette contrainte inhérente à toute mesure, et par voie de conséquence à toute démarche d'évaluation, constitue fort probablement une dernière enclave dont l'évaluation formative ne peut être facilement libérée.

On voit donc que tout est en place pour une libération de l'évaluation formative des principaux éléments qui lui ont été associés au départ. Elle peut être au service de pédagogies autres que la pédagogie de maîtrise, sans pour autant renier toute velléité de donner à chaque élève les meilleures chances de réussir. La pratique de l'évaluation formative doit aussi être orientée vers des préoccupations beaucoup plus didactiques que docimologiques. Enfin, il reste à développer des stratégies de collecte d'observations qui permettent de retracer des processus, des cheminements empruntés par des élèves et ce, à partir d'indicateurs «de surface» qu'il faudra traiter de façon particulière (voir le texte de Jean Retschitzki dans le présent ouvrage et celui de Scallon (1988b) qui contient un chapitre sur l'observation des erreurs systématiques en évaluation formative). Il apparaît de plus en plus évident que cette libération correspond étroitement à une préoccupation dominante : celle de fournir à chaque élève (et aussi aux enseignantes et aux enseignants) un feed-back de qualité. Il reste à démontrer que le développement récent des sciences cognitives vient renforcer cette préoccupation et ouvre des horizons tout à fait nouveaux pour la pratique de l'évaluation formative.

4. LA RÉGULATION : PAS AUSSI SIMPLE QUE L'ŒUF DE COLOMB !

Dans les premiers écrits reliés de près ou de loin à la pratique de l'évaluation formative, les correctifs à prescrire au sujet d'un rendement inadéquat n'ont pas reçu d'étiquette particulière. Le plus souvent, il s'agissait d'assigner aux élèves visés des exercices additionnels de pratique ou de leur fournir des explications en reprise de l'enseignement initial. Block et Anderson (1975) ont décrit toute une panoplie de procédés dits correctifs qui se distinguent selon que l'on mise sur la cognition ou la motivation, sur une approche individualisée ou un travail de groupe, ou encore sur des modalités de présentation du matériel.

Qu'en est-il du mot de plus en plus rencontré de nos jours pour traiter de la visée essentielle de l'évaluation formative : la *régulation* ? Il est assez bizarre de constater que certains répertoires tels le *Current Index of Journals in Education (C.I.J.E.)* ne proposent aucun mot clé au titre de *régulation* et que les références à l'enseignement correctif (« corrective teaching », « remedial teaching ») visent habituellement des programmes entiers de traitement pour des groupes d'élèves présentant des caractéristiques particulières (p. ex., élèves sous-doués, élèves handicapés, troubles de comportement, etc.) ce qui nous éloigne considérablement des limites dans lesquelles il faut situer l'évaluation formative des apprentissages.

Dans le même ordre d'idées, il est relativement facile de constater que la pratique même de l'évaluation formative en milieu scolaire a connu des dérapages importants. C'est ce qui se dégage d'un rapport du Conseil supérieur de l'éducation du Québec dans lequel l'évaluation des apprentissages au primaire fait l'objet d'un tour d'horizon (1992). Il serait laborieux de décrire tous les aspects traités dans ce document mais il apparaît de toute évidence, pour ce qui est de la pratique de l'évaluation formative telle que signalée dans le rapport, que la notion de régulation n'est pas toujours à l'avant plan des préoccupations du milieu et il n'est même pas certain qu'elle soit bien comprise.

Le terme « régulation » semble émerger de certains écrits européens, puisqu'on le trouve au centre de travaux de réflexion sur l'évaluation formative notamment ceux de Allal (1978) et de Cardinet (1986). La définition formelle de la régulation, trouvée dans un dictionnaire, se lit comme suit à propos d'un dispositif ou d'un mécanisme :

> Fait d'en régler le fonctionnement ou le mode de fonctionnement, notamment pour l'adapter aux conditions extérieures ou au résultat à obtenir (CNRS, 1990).

Cette définition n'ajoute rien à la compréhension actuelle de ce que doit être la régulation. Les synonymes qui s'y rapportent sont tout de même intéressants et devraient aider à se saisir de cette notion importante : *régler*, *adapter*, *ajuster*, *contrôler*, ou *mettre au point*..

À mon humble avis, la notion de régulation mérite d'être bien mise en évidence pour aborder la façon de concevoir la pratique de l'évaluation formative dans des contextes pédagogiques autres que ceux qui lui ont donné naissance. Lorsque la question de l'évaluation formative est posée on doit maintenant se tourner vers une question beaucoup plus précise : *La régulation : c'est l'ajustement de qui ou de quoi ?* De l'élève pour que celui-ci s'accorde à un ensemble d'objectifs bien planifiés ? Des actions pédagogiques et didactiques pour mieux les adapter aux caractéristiques personnelles des élèves ? Tel est pour l'instant l'enjeu qui se présente à toute clarification des visées de l'évaluation formative en milieu scolaire.

L'auteur de ce texte a voulu tester la définition donnée ci-haut de la régulation en essayant de visualiser un épisode typique d'une pédagogie de situation inspirée de travaux tout récents en didactique de la lecture (Demers et Tremblay, 1992; Tremblay et Scallon, 1994). Il s'agit essentiellement d'une pédagogie qui offre aux élèves des moments de lecture libre, des moments de lecture guidée et des projets de lecture autonome. C'est à l'occasion de projets de lecture autonome que les élèves sont sollicités pour une démarche de vérification de leur capacité de lecture. Ils sont d'abord saisis d'une tâche à réaliser (p. ex., apprendre des règles pour réaliser un jeu, démarrer une collection de pierres, compléter une fiche sur un animal, etc.). Chaque projet comporte une mise en situation à laquelle tous les élèves de la classe participent simultanément, une période de lecture autonome (travail individualisé) et un retour collectif sur le parcours de lecture, sur la réussite de la tâche et sur l'expérience vécue par les élèves. L'approche n'est pas exclusive à la didactique de la lecture car l'apprentissage par problèmes en médecine présente une mise en scène comparable : saisie d'un cas clinique en petit groupe en présence d'un professeur, recherche d'informations pertinentes et recours à des stratégies appropriées sur une base individuelle, retour en groupe pour analyser le diagnostic à poser et échanger sur les difficultés de parcours. Berkson (1990) a présenté une description très élaborée de ce qui se présentait au départ comme une approche pédagogique et qui est devenue un programme de formation en médecine.

Pour rendre compte des divers moments de régulation dans un épisode de pédagogie de situation, la figure 1 présente les ajustements réalisables depuis le point de départ (la saisie du projet à réaliser) jusqu'au retour réflexif. Le corps de cette figure rappelle que l'accomplissement de la tâche dépend de la maîtrise de plusieurs éléments empruntés au vocabulaire des sciences cognitives : connaissances déclaratives, connaissances procédurales, stratégies, habiletés (skills), etc. On comprendra toutefois que la maîtrise de ces éléments n'est pas exigée avant d'aborder un projet, comme en pédagogie de maîtrise. On ne peut cependant nier leur importance pour réaliser la tâche avec succès. C'est en suivant le parcours présenté dans la figure 1 que l'on peut mettre en évidence toute l'étendue du registre des régulations qui s'offrent à une pédagogie de situation :

- *au départ :* le projet doit correspondre au niveau de compétence et aux caractéristiques des élèves en général; la clarté des consignes, certains aspects du projet et les explications fournies doivent être objets de réglage et ce, dès le début du projet.

  Le bien fondé du projet doit être expliqué pour susciter la motivation des élèves; les difficultés à surmonter doivent être perçues de façon telles que les élèves éprouvent un sentiment de «contrôlabilité» de la tâche; ici encore plusieurs réglages sont possibles et peuvent même être anticipés lors de la planification. Cette phase est cruciale car c'est bien la tâche à réaliser qui donne aux élèves une intention de lecture dans ce type de pédagogie. De plus, c'est le résultat à obtenir ou le but à atteindre qui doit donner emprise à la régulation ou à l'autorégulation.

- *en cours de route :* connaissances, habiletés, stratégies sollicitées d'un projet à l'autre peuvent faire défaut; l'ajustement ici peut être envisagé en terme de consolidation par la pratique au moyen d'exercices structurés à visée didactique (feed-back); par exemple, une difficulté éprouvée par plusieurs élèves déclenchera le recours à une démarche instrumentée pour vérifier de façon formelle l'utilisation adéquate de certaines stratégies

  Lecture et réalisation de la tâche sur le plan individuel ; *on est au cœur du projet où l'autonomie de chaque élève est le plus sollicitée* : surveillance étroite par l'élève de son cheminement (monitoring) et remise en question en fonction du résultat à obtenir (régulation menée après autoévaluation ou activité de métacognition).

- *au retour :* rétroaction sur le parcours suivi par chaque élève, sur l'utilisation de ses propres ressources et de celles mises à sa disposition et rétroaction sur la réussite du projet; ajustements à long terme en préparation de projets à venir. La mise en scène en est une basée sur des échanges en classe.

La lectrice et le lecteur intéressés trouveront une présentation beaucoup plus détaillée et nuancée dans un ouvrage récent paru au regard de cette didactique de la lecture (voir Tremblay et Scallon, 1994).

La démarche qui vient d'être décrite fait voir assez clairement le double objet de la régulation et le rôle des acteurs en présence. D'une part, on voit que la situation pédagogique, c'est-à-dire le projet et ses éléments, est objet de régulation dès le départ. D'autre part, la démarche de l'élève avec ses connaissances, ses habiletés et sa motivation est aussi traitée comme un objet important de régulation. Enfin, il serait superflu d'insister sur le rôle de l'enseignant et celui des élèves dans la mise en œuvre des multiples actions d'évaluation et de régulation.

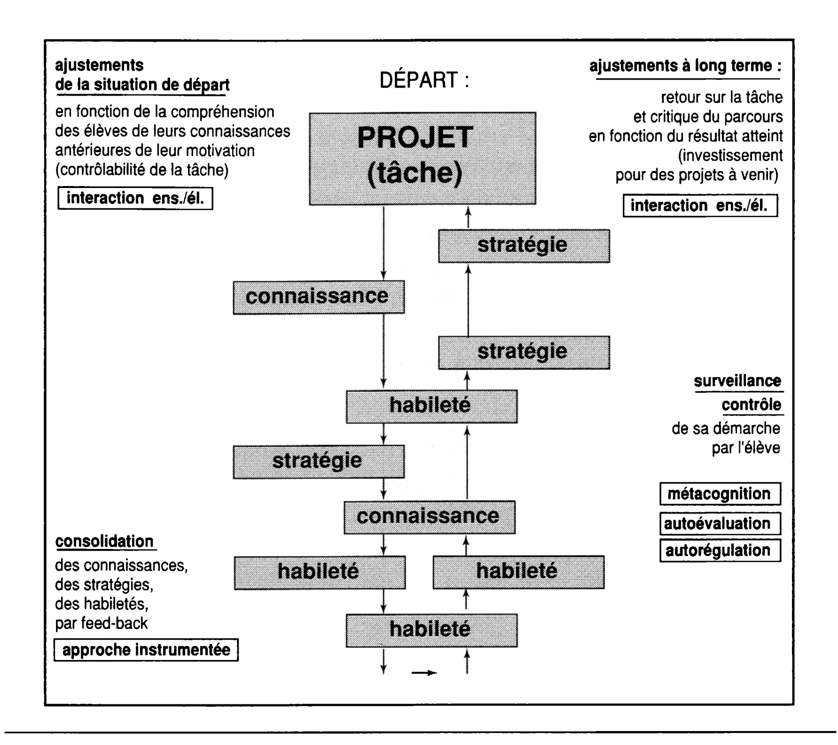


FIGURE 1

Événements de régulation dans une pédagogie de situation

5. L'APPORT DES SCIENCES COGNITIVES : QUELS MATÉRIAUX NOUVEAUX POUR L'ÉVALUATION FORMATIVE ?

Situer les objets de réglage, d'ajustement ou de régulation, dans une pédagogie aux parcours encore inédits, inspirée des théories de l'apprentissage du dernier cri, est une chose. Savoir sur quels indices visibles jeter son regard, sur quelles bribes d'information fonder une décision ou encore dans quelle direction précise diriger son attention en est une toute autre. Et les enseignants doivent savoir et les élèves aussi car, d'après ce qui vient d'être écrit, ce sont les principaux acteurs qui sont engagés dans la pratique de l'évaluation formative.

Jusqu'ici les apprentissages cibles de l'évaluation formative ne se sont pas tellement différenciés de ceux habituellement visés par l'évaluation en général : soit des résultats à atteindre exprimés le plus clairement possible sous forme d'objectifs d'apprentissage. Des taxonomies et diverses tentatives de classification se sont accumulées au fil des ans comme en faisait foi un bilan mené en 1975 par De Landsheere et De Landsheere. Mais aucune de ces clas-

sifications ne semble avoir influencé les pratiques d'évaluation au point d'imposer des subtilités de nature méthodologique qui puisse différencier nettement l'évaluation formative d'autres fonctions de l'évaluation.

Mais qu'en est-il des propositions venant du domaine des sciences cognitives. Des écrits récents faisant état des problèmes relativement nouveaux qui se posent à l'appréciation d'habiletés cognitives sont très indicateurs à ce sujet. Par exemple, au-delà de la simple mesure des connaissances au moyen de techniques de rappel (i.-e. avec des tâches de restitution de réponses apprises par exemple) on semble s'intéresser de plus en plus à leur structure ou à leur organisation chez les élèves (Royer, Cisero et Carlo, 1993). Dans le cas de procédures à faire suivre par des sujets, on peut s'intéresser non seulement à l'exactitude de leur exécution mais aussi au degré d'automaticité avec lequel elles ont été effectuées (Anderson, 1982). Et la liste des éléments nouveaux, voire inédits, pourrait s'allonger davantage. La figure 2 fournit un aperçu des divers éléments empruntés au domaine des sciences cognitives et avec lesquels il faut composer pour ajuster les pratiques d'évaluation des apprentissages. On remarquera que le domaine est encore en friche et qu'il ne semble pas y avoir de principe unificateur qui puisse orienter dans une direction commune tous les investissements méthodologiques à venir.

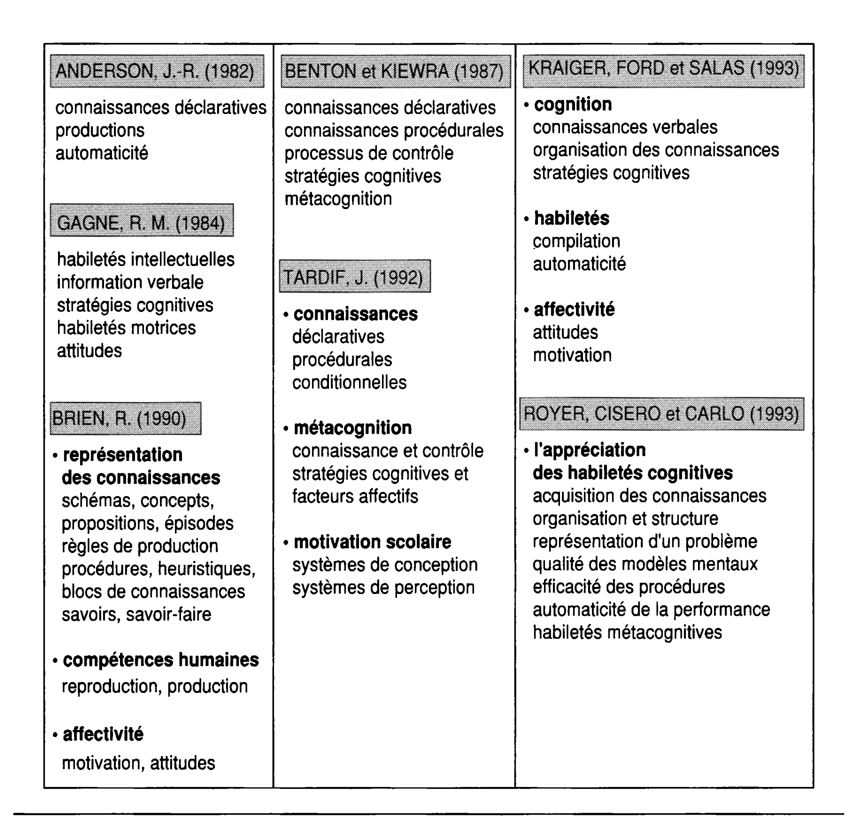


FIGURE 2

Aperçu des mots clés utilisés dans quelques ouvrages récents se rapportant au domaine des sciences cognitives

Il n'est pas dans le cadre de ce texte de reprendre une à une les composantes qui pourraient servir de cibles à l'évaluation formative et d'en traiter avec profondeur. Tout au plus pouvons nous envisager d'une manière purement hypothétique ce que l'avenir nous réserve. Pour y arriver, certaines questions doivent être élucidées.

La très grande majorité des textes qui ont été consultés au regard des procédés d'observation et d'appréciation (assessment) inspirés par les sciences cognitives s'adressent beaucoup plus à l'évaluation pronostique ou diagnostique qu'à l'évaluation formative. En complément à cette perspective, les recencions d'écrits de nature méthodologique ont plutôt fait état des procédés utilisés pour évaluer des programmes d'intervention ou des moyens d'enseignement (p. ex., Benton et Kiewra, 1987; Kraiger, Ford et Salas, 1993; Royer, Cisero et Carlo, 1993). Pour ce qui est de la pratique de l'évaluation formative, la nuance est importante, tant du point de vue méthodologique et du point de vue des actions à poser. Beaucoup de caractéristiques d'intérêt, telles la structure des connaissances ou des concepts par exemple, exigent des procédés élaborés de cueillette d'informations et de traitement de données qui ne sont certes pas à la portée des enseignantes et des enseignants au contact d'une classe d'élèves. C'est ainsi que se présente le degré d'organisation des connaissances dans la structure cognitive d'un élève, l'un des cas les mieux documentés dans les écrits. De plus, les possibilités de régulation des divers phénomènes d'intérêt sont pour la plupart méconnues et les écrits en évaluation diagnostique n'ont pas cette préoccupation. On voit donc tout le problème qui doit se poser à la formation des maîtres lorsqu'il s'agira de leur proposer des cibles nouvelles pour l'évaluation formative, cibles empruntées à un domaine encore en émergence comme celui des sciences cognitives.

6. LA LUMIÈRE AU BOUT DU TUNNEL : LA MÉTACOGNITION

Pour l'instant, d'après l'état actuel de nos connaissances, il n'est pas certain que les diverses habiletés et les stratégies cognitives aient été «pédagogisées» à ce point qu'elles puissent être apprises ou enseignées pour être traitées éventuellement de simples objets de régulation. Par exemple, Romainville (1993) a mis en doute d'une certaine manière la possibilité d'entraîner des sujets à utiliser des stratégies d'étude (p. 38 et suivantes). La critique s'applique-t-elle également à tout un registre d'habiletés diverses ? Les textes recensés ne sont pas explicites sur ce point mais on peut l'inférer facilement dans le cas de l'ouvrage de Tardif (1992) traitant plus particulièrement de l'enseignement de stratégies. Cependant, nombre de variables d'intérêt qui alimentent le discours des sciences cognitives sont actuellement traitées le plus souvent dans le cadre de recherches exploratoires ou dans celui de l'attestation de l'efficacité de programmes d'intervention auprès de clientèles à problème ou en difficulté. Si l'importance à accorder à ces variables (on peut

penser ici au sentiment d'efficacité personnelle, ou à l'attribution causale par exemple) faisait en sorte qu'elles doivent se situer en amont de tout programme de formation, elles seraient alors la matière première de l'évaluation diagnostique beaucoup plus que de l'évaluation formative. Or, «l'éducabilité» de certaines de ces caractéristiques commence à se manifester à l'horizon de nos préoccupations. C'est le cas notamment des changements apportés au système d'attribution causale des individus en relation avec une structure de feedback ou de notation plus personnalisée, voire dans certains cas, critérielle (Oren, 1983; Covington et Omelich, 1984; McColskey et Leary, 1985; Butler et Nisan, 1986). Tardif (1992) rapporte plusieurs études qui auraient obtenu des effets positifs pour l'enseignement de stratégies de résolution de problèmes (pp. 251 et suiv.). Ce ne sont là, bien sûr, que quelques pistes de réflexion et il faudrait consacrer beaucoup plus de lignes que les limites de ce texte n'autorisent.

De tous les éléments mis en évidence ces dernières années il en est un qui semble persister et même résister à tous les courants : il s'agit de l'autoévaluation. Si certaines personnes ont cru assister à l'émergence d'une mode ils peuvent maintenant modifier leur perception car c'est devenu un pôle d'attraction fort documenté et il serait difficile pour ne pas dire presque impossible de rendre compte de l'abondance des écrits qui en ont traité. Dans le cadre de ce texte, qu'il soit dit que l'autoévaluation se conçoit fort bien comme une modalité d'appropriation de l'évaluation formative mais aussi (et certains diraient d'abord et avant tout) comme un outil indispensable à toute entreprise de formation. C'est ainsi qu'il fallait interpréter la conception développée par certains chercheurs d'Aix-en-Provence au milieu des années'80 et, lors d'une rencontre qui m'a été des plus profitables, je me suis permis d'utiliser l'étiquette d'«évaluation formatrice» pour cerner une conception de l'évaluation formative fort distincte de celle des pionniers. L'idée de faire de l'évaluation formative une habileté à développer chez les élèves, un savoir être, a été fort bien élaborée et soutenue par Nunziati (1990).

Dans son évolution comme concept, au travers des années, l'autoévaluation est apparue de plus en plus comme un fil directeur et un lieu de nécessité qui en a imposé de plus en plus à la méthodologie de l'évaluation formative. En effet, nul ne saurait aujourd'hui se contenter d'une approche globale et superficielle qui ferait de l'autoévaluation une simple affaire d'impression générale, ou un état d'âme dont l'individu est plus ou moins conscient. Dans le cadre de certaines pédagogies de situation on l'a vue être greffée à des productions complexes mais concrètes telles qu'il s'en trouve dans le domaine de l'enseignement professionnel (par exemple, au Québec) ou encore en français langue maternelle. Le concept «d'objectivation des pratiques de communication», tel qu'entendu dans les programmes de français (primaire et secondaire) du Ministère de l'éducation au Québec, était et est encore nettement de cet ordre. Cependant, pour écourter un peu l'histoire et l'évolution des idées, la notion d'autoévaluation semble avoir été associée au départ à celle de production élaborée que l'élève doit réaliser dans un contex-

te de relative autonomie (cf. Leselbaum, 1982). Qu'en est-il de l'autoévaluation par l'élève de ses connaissances sans passer par un jugement impressionniste ? Qu'en est-il de l'autoévaluation de certaines habiletés cognitives que l'on peut cerner avec des épreuves à correction objective ? Il faut reconnaître que, pendant plusieurs années, le discours et la pratique de l'autoévaluation sont demeurés assez imperméables à ces réalités diverses.

La notion de «métacognition» est venu jeter un éclairage tout à fait nouveau au regard de toutes ces questions. Savoir que l'on comprend, que l'on sait, que l'on sait faire, apparaît de plus en plus être à la base de l'autoévaluation. S'évaluer (ou s'autoévaluer) au sujet d'une prestation en tenant compte de plusieurs points de vue ou de plusieurs critères est une chose; s'évaluer dans le but de poser des gestes de nature corrective, dans le but de s'ajuster (et on parle ici de régulation faite par l'individu lui-même c'est-à-dire d'autorégulation) en est une toute autre. Une telle évaluation doit être réaliste ou encore menée avec «justesse». C'est ici, énoncé un peu brutalement, que la métacognition entre fort probablement en jeu puisque s'autoévaluer avec justesse c'est précisément savoir que l'on sait (quand on sait) ou que l'on ne sait pas (quand on ne sait pas). C'est sans doute ainsi que Allal (1993) traite du processus de régulation métacognitive et de la place qu'il devrait occuper dans l'évaluation formative par l'élève. L'auteur propose un canevas qui pourrait s'avérer utile pour articuler une démarche d'autoévaluation : une phase d'anticipation constituée d'activités de prévision pour orienter l'action, une phase de contrôle consistant en un processus continu de comparaison entre un état donné et un résultat à atteindre (monitoring) et une phase d'ajustement (la régulation proprement dite). Dans ce paradigme il revient à l'évaluation formative de soutenir divers niveaux d'autorégulation (Allal, 1993, pp. 88-89).

Il est un autre aspect intéressant à souligner lorsqu'on situe la métacognition au coeur d'une pratique d'évaluation formative qui fait de celle-ci une habileté fondamentale à développer chez les élèves. Il apparaît de plus en plus, d'après l'opinion explicite de certains chercheurs, que la métacognition peut être «éducable» ou encore «pédagogisable» au-delà du seul fait qu'elle se développe avec l'âge. Chez quelques auteurs qui ont attiré l'attention sur ce point, on parle de «conscience métacognitive» (Noël, 1991) ou de «prise de conscience de son fonctionnement» (Romainville, 1993). Cette conscience peut être objet d'entraînement ou d'action pédagogique et Romainville, dans cette perspective, a beaucoup insisté sur l'importance du feed-back ou information en retour (1993, pp. 59-60).

Pour récapituler, en bref ...

Tout en maintenant le cap sur la régulation de l'apprentissage, les conceptrices et les concepteurs du discours sur l'évaluation formative nous ont proposé des savoir-faire variés qui reflètent assez bien diverses conceptions de la pédagogie et aussi diverses représentations de ce qui doit être objet d'ajustements ou de réglages. On a vu assez rapidement que les premières ver-

sions méthodologiques devaient répondre aux impératifs d'une pédagogie centrée sur des objectifs bien circonscrits avec en plus le souci de démocratiser le plus possible la maîtrise de ces objectifs. L'avènement de pédagogies plus complexes, centrées sur des objets tout aussi complexes (apprentissage par problème ou pédagogie de situation, par exemple) ont quelque peu ébranlé une méthodologie qui n'a pu se poser comme tradition. C'est dans cette dynamique de changement constant de définition et d'orientation que le domaine des sciences cognitives a pu influencer et influencera davantage la pratique de l'évaluation formative. Non pas que ce type d'évaluation devienne altéré conceptuellement !

C'est plutôt dans l'ordre de ses appropriations qu'il faut diriger toute notre attention. L'autoévaluation, la prise en charge de ses apprentissages et de sa progression par l'élève et au mieux la possibilité de poser elle-même ou lui même ses propres actions de régulation est sans contredit l'appropriation la plus marquée. Si bien des caractéristiques relativement nouvelles émergent du discours des sciences cognitives (connaissances déclaratives, procédurales, conditionnelles, etc.) la métacognition se présente déjà comme un centre d'intérêt tant pour la recherche des prochaines années que pour la pratique pédagogique. L'évaluation formative ne pourra que bénéficier de la jonction entre le traitement plus approfondi de cette conscience de ses processus cognitifs ou de ses démarches de la part de l'élève, de l'étudiant, d'une part, et les phénomènes d'autoévaluation et d'autorégulation qui ont été mis en évidence depuis plusieurs années.



Analyse des pratiques éducatives visant à faire participer l'apprenant à l'évaluation diagnostique, au pilotage et à la régulation de ses apprentissages

José-Luis Wolfs

Introduction

Les pratiques éducatives visant à amener les élèves à piloter, auto-évaluer et réguler leurs apprentissages se sont fortement développées ces dernières années. L'objectif de ce texte est de proposer un instrument permettant d'analyser ces pratiques, en référence essentiellement aux travaux relatifs à la métacognition.

1. LA PARTICIPATION DE L'ÉLÈVE À L'ÉVALUATION DIAGNOSTIQUE, AU PILOTAGE ET À LA RÉGULATION DE SES APPRENTISSAGES

Les pratiques pédagogiques qui visent à aider l'élève à participer de manière aussi active et autonome que possible au pilotage, à l'évaluation et à la régulation de ses apprentissages se sont fortement développées ces dernières années, ainsi qu'en témoignent :

- les travaux des chercheurs en Sciences de l'Education (ex : Allal, 1988, Allal, 1993, Bonniol, 1981, De Ketele, 1991, Laveault, 1990, Leclercq, 1991, Paquay, Allal, Laveault, 1990...);
- les recommandations officielles émanant de pouvoirs organisateurs d'enseignement (ex : une publication de la Commission de Rénovation de l'Enseignement fondamental consacrée à l'évaluation formative et à l'auto-évaluation, Ministère de l'Education en Belgique francophone, à paraître prochainement);
- les ouvrages consacrés à ce thème et destiné aux enseignants (Leselbaum, 1982, De Vecchi, 1992...).

Il apparaît non seulement que ces pratiques sont extrêmement diversifiées, mais qu'en outre les mêmes termes recouvrent parfois des réalités très différentes. Ainsi, comme le notait De Ketele, dans un article intitulé avec humour *«Ballade au pays de l'auto-évaluation»*, certaines pratiques se réclamant de l'auto-évaluation conduisent en fait à la dépendance alors que d'autres favorisent l'autonomie.

C'est pourquoi, il nous paraît judicieux de tenter de proposer un instrument d'analyse des différentes pratiques visant - au sens large - à aider les élèves à piloter, auto-évaluer et réguler leurs propres apprentissages. Avant de développer les différents critères que nous suggérons (dans la troisième partie de cet article), nous nous efforcerons de clarifier et de systématiser le concept de métacognition, qui sous-tend, explicitement ou non, la plupart d'entre elles.

2. LA MÉTACOGNITION

2.1 Origine du concept

L'émergence et le développement du concept résultent de la convergence de plusieurs courants : recherches de Flavell (1976) sur la mémoire et la métamémoire, de Piaget (1974) sur la prise de conscience, de Vygotsky sur les origines sociales du contrôle cognitif, de Sternberg sur le processus de contrôle dans le traitement de l'information.

Toutefois, c'est Flavell (1979) qui a le plus contribué à populariser le concept de métacognition. Sa définition englobait deux composantes *(«knowledge and cognition about cognitive phenomen» / «cognitive monitoring»)* qui ont été reprises par la plupart des auteurs ultérieurs.

Gombert (1990), en particulier définit la métacognition comme suit : *«Domaine qui regroupe : (1) les connaissances introspectives conscientes qu'un individu particulier a de ses propres états et processus cognitifs, (2) les capacités que cet individu a de délibérément contrôler et planifier ses propres processus cognitifs en vue de la réalisation d'un but ou d'un objectif déterminé.»*

Dans un souci de rendre opérationnel le concept, la définition de Gombert a le mérite, comme le suggéraient déjà Kitchener (1983) et Noël (1988), d'insister sur le fait que les connaissances métacognitives sont introspectives et portent sur le fonctionnement cognitif du sujet lui-même. Elles doivent être clairement distinguées des connaissances d'ordre général relatives aux phénomènes cognitifs (psychologie, épistémologie...) que le sujet pourrait en outre avoir. Ces dernières peuvent bien entendu devenir métacognitives à leur tour, si le sujet les utilise pour analyser son propre fonctionnement cognitif.

Nous tenterons à présent, en nous référant essentiellement à un modèle élaboré en commun avec Noël et Romainville (Noël, Romainville, & Wolfs, à paraître) de décrire schématiquement ses différentes modalités de fonctionnement en distinguant bien les deux dimensions retenues par la plupart des auteurs : connaissances métacognitives et contrôle métacognitif (ou régulation), car comme nous le montrerons plus loin, ces deux dimensions ne sont pas toujours liées.

En ce qui concerne le problème des régulations, signalons un intéressant article de Allal (1993), où l'auteur tente d'articuler les différents types de régulations dans une perspective systémique : régulations d'ordre institutionnelles, régulations entre l'enseignant et l'élève ou entre élèves et régulations métacognitives (propres au sujet lui-même).

2.2 Modalités de fonctionnement de la métacognition

Nous distinguerons schématiquement trois grands types de situations que nous développerons ci-après :

- situation 1 : régulation de l'activité cognitive sans intervention d'une réflexion métacognitive;
- situation 2 : développement d'une réflexion métacognitive;
- situation 3 : réflexion métacognitive et régulation.

2.2.1 *Situation 1 : régulation de l'activité cognitive sans intervention d'une réflexion métacognitive*

Nous pouvons distinguer deux cas :

CAS 1 : Le sujet (ici l'apprenant) contrôle ou régule ses activités cognitives en utilisant des processus implicites (dans le sens de Dulany, 1984) ou des processus automatisés, sans intervention d'une réflexion métacognitive.

Exemple de processus automatisés (cité par Gombert, 1990) :

« Tout d'abord, l'acte graphique lui-même, le choix des mots, l'orthographe, les accords grammaticaux, la ponctuation, sont souvent opérés automatiquement. De même, la recherche de cohérence et de cohésion textuelle, l'adaptation au lecteur potentiel ne font pas toujours l'objet d'une attention particulière alors que le texte sera satisfaisant sur ces critères. Enfin, Bartlett (1982) signale que de nombreuses corrections sont effectuées si rapidement que les sujets n'ont pas pu réfléchir sur la nature du problème. Tout porte à penser qu'il s'agit là de processus automatisés, cette automatisation permettant que la plus grande part de l'effort cognitif soit consacré à l'élaboration du contenu. Il est toutefois notable que chacune de ces activités peut être contrôlée délibérément si la situation l'impose ou si un obstacle survient dans la rédaction. Ceci suggère que, dans l'apprentissage, l'automatisation a été consécutive à une maîtrise consciente puis à un exercice de ces différents composants. »

D'une manière plus générale, une situation de nouveauté, de difficulté, de « déséquilibre cognitif » serait propice à l'exercice d'une réflexion métacognitive. A l'inverse, une tâche familière et répétitive conduirait à l'automatisation progressive de « routines » de comportements.

CAS 2 : Les activités cognitives du sujet sont guidées par une procédure, une marche à suivre fournie de l'extérieur.

Exemple : C'est le cas de certaines grilles qui prescrivent à l'apprenant une marche à suivre très précise, sans l'inviter à expliciter, analyser ou évaluer sa démarche.

Dans les cas 1 et 2, il apparaît clairement un contrôle et une régulation des activités cognitives ou des apprentissages, sans toutefois qu'intervienne une réflexion métacognitive, si ce n'est incidemment.

2.2.2 *Situation 2 : le sujet développe une réflexion métacognitive*

Nous pouvons distinguer - schématiquement - plusieurs activités métacognitives : *l'explicitation* (évocation, description par le sujet de ses processus cognitifs), *l'analyse* (mise en relation entre plusieurs processus, entre processus et variables contextuelles, entre processus et résultats...), *la conceptualisation* : abstraire de différentes situations analysées des propriétés générales relatives à son propre fonctionnement, des règles d'action ou stratégies applicables à différents contextes.

Ces différentes activités métacognitives d'un sujet peuvent porter sur différents objets : *son propre fonctionnement cognitif* (passé, présent ou futur) d'une part ou les *variables externes* qui influencent celui-ci d'autre part. Enfin, ces activités peuvent se dérouler selon un *mode descriptif* ou un *mode évaluatif*. Dans ce dernier cas, le sujet émet un jugement sur la qualité de son fonctionnement cognitif.

Pour illustrer ces différentes formes de réflexion métacognitive, nous citerons quelques exemples de réponses fournies par des étudiants de première année d'études universitaires à deux enquêtes : l'une portant sur les comportements d'étude et de préparation aux examens en général (Fonteyn, 1993) et l'autre portant sur la prise de notes en particulier (Lona, 1994).

Illustration :

(1) L'explicitation :

Explicitation / mode descriptif / fonctionnement personnel

- «Dans ma prise de notes, j'écoutais d'abord la phrase entière avant d'écrire...»
- «... en abrégeant un maximum de mots»
- «... j'ai privilégié les flèches»

Explicitation / mode descriptif / variables externes

- «J'ai surtout tenu compte de ce que le professeur projetait à l'écran»
- «... sur lesquelles le professeur a insisté»
- «... vu la lecture rapide»

Explicitation / mode évaluatif / fonctionnement personnel

- «Je me suis trop basée sur le mot-à-mot.»
- «... c'est comme ça que je fais et ça m'a réussi»

Explicitation / mode évaluatif / variables externes

- «Le professeur ne facilitait pas la prise de notes.»

(2) L'analyse

L'étudiant exprime des relations (cause, conséquence, temps, but/avantage, moyen, condition, alternative, opposition, restriction...) entre processus cognitifs, entre ceux-ci et des variables contextuelles, entre processus et résultats... (Pour plus d'informations sur l'activité d'analyse, voir Romainville, 1992, 1993.)

Analyse / Mode descriptif / Fonctionnement personnel

- «Ne pas étudier quelque chose d'incompris car on ne le retient pas.» (Conséquence).
- «Le fait de relire le cours et ses notes permet de voir si effectivement tout est bien compris.» (But/avantage)
- «Reformuler autrement s'il y a incompréhension» (Condition)
- «Assister au cours... cela pallie le manque d'étude régulière du syllabus.» (Alternative)

Analyse / Mode descriptif / Variables externes

- «... lorsqu'il m'arrive de ne pas avoir le temps d'écrire mes phrases.» (Condition)
- «Structurer le cours (...) Attention toutefois à ne pas donner trop d'importance au support matériel.» (Restriction)
- «... car c'était un travail à rendre et non des notes personnelles» (Opposition)

Analyse / Mode évaluatif / Fonctionnement personnel

- «Recopier ses notes : j'ai une mémoire visuelle, donc je mémorise très facilement de la sorte». (Cause)
- «En faisant ressortir les points essentiels de la sorte, j'ai déjà une idée assez précise du contenu lorsque je reverrai mes notes par la suite.» (Moyen)
- «La raison pour laquelle j'essaie de noter le plus possible est que je ne retiens pas longtemps les mots entendus.» (Cause)

Analyse / Mode évaluatif / Variables externes

- «... Il est impossible d'ingurgiter la matière telle qu'elle est présentée dans le syllabus.» (Opposition)
- «Ceci n'était pas difficile, dans la mesure où cette activité se faisait entre copains, dans un climat de détente.» (Condition)

(3) La conceptualisation

Ici aussi, nous pourrions proposer - en principe - des exemples illustrant les différentes sous-modalités de cette activité : mode descriptif/évaluatif, fonctionnement personnel/variables externes. Toutefois, nous nous en abstiendrons provisoirement, dans la mesure où il est

extrêmement difficile à partir d'extraits de réponse cités hors contexte, de savoir s'il y a eu effectivement conceptualisation ou non.

Considérons les extraits suivants :

- «Il est toujours plus facile d'étudier ses propres notes.»
- «On se rappelle mieux les choses contre lesquelles on a des griefs, auxquelles l'on a réfléchi.» (Cette réflexion d'un étudiant rappelle une des pensées de PASCAL qui disait - en substance - : «On se persuade davantage des raisons que l'on a soi-même trouvées que de celles qui nous ont été données.»)

Tous ces énoncés possèdent un haut degré de généralité, toutefois, il est difficile de savoir dans quelle mesure ils sont le fruit d'une abstraction portant sur de nombreuses expériences personnelles dûment analysées ou si l'étudiant reprend simplement à son compte une croyance, un lieu-commun. (Ex : «Il y des méthodes clés pour réussir à l'Université. Le plan, le résumé, la logique, la synthèse font partie de ces méthodes.»)

C'est pourquoi, vu la difficulté d'opérationaliser les activités de conceptualisation, nous nous limiterons dans notre analyse des outils pédagogiques proposés aux élèves, à distinguer les questions qui renvoient à un contexte spécifique bien défini («Comment as-tu fait hier soir pour étudier ta leçon de géographie ?») et celles qui sont décontextualisées («Comment fais-tu habituellement pour étudier tes leçons ?»).

2.2.3 *Situation 3 : Réflexion métacognitive et régulation*

Le sujet utilise sa réflexion et ses connaissances métacognitives pour contrôler et réguler ses activités. En fonction des résultats obtenus, certaines de ses connaissances ou stratégies métacognitives se voient renforcées (validées) ou au contraire remises en question et modifiées.

Pour schématiser encore plus ces différentes modalités de fonctionnement, nous pourrions les résumer comme suit :

1. le sujet se trouve en situation de *"pilotage automatique"* (situation 1, cas 1) :
2. le sujet se trouve en situation de *"pilotage externe"* (situation 1, cas 2);
3. le sujet se trouve en situation de *"pilotage interne"* à des degrés divers

   situations 2 et 3).

Le problème de la relation entre «réflexion métacognitive» et régulation» est complexe. L'élève peut en effet réguler certaines de ses activités sans qu'intervienne une réflexion métacognitive et à l'inverse disposer de

connaissances métacognitives et ne pas les utiliser. Nous avons d'ailleurs consacré un article antérieur à l'examen de ce problème (Wolfs, 1992).

Nous nous limiterons ici à citer Glaser (1986) : «Le rôle de ces habiletés (relatives aux capacités de contrôle, à la métacognition) serait relativement limité dans les démarches intellectuelles s'appliquant à des domaines de connaissances spécifiques où le sujet dispose facilement de schèmes ou de procédures qui lui permettent de traiter rapidement les données d'un problème». Toutefois, il ajoute : «On peut enseigner à des étudiants la connaissance d'une règle, d'une théorie ou d'une procédure, mais si on prend pour critère le transfert des apprentissages à de nouvelles situations, il faut en outre que ces étudiants sachent contrôler eux-mêmes l'utilisation de leurs connaissances.»

Ce rôle de la métacognition comme *agent facilitateur du transfert d'apprentissage* a été particulièrement mis en évidence par Sternberg et Wagner (1984), Derry, & Murphy (1986) au terme d'une large revue de la littérature portant sur l'étude des facteurs exerçant une influence sur l'efficacité (le transfert) des programmes visant le développement des «thinking skills» et «learning skills».

3. ANALYSE DES DIFFÉRENTES PRATIQUES ÉDUCATIVES visant à aider les élèves à piloter, auto-évaluer et réguler leurs propres apprentissages

Nous nous centrerons essentiellement sur les pratiques qui utilisent comme support des questionnaires ou des grilles visant - au sens large - à susciter chez l'élève une réflexion sur ses apprentissages (passés, présents ou futurs) et/ou un pilotage et une régulation de ceux-ci. Nous proposons ci-après une grille d'analyse de ces pratiques éducatives visant à identifier :

- le (ou les) objectif(s) pédagogique(s) de l'outil proposé aux élèves (critère 1);
- les facettes du fonctionnement cognitif et/ou métacognitif concernées par le questionnaire ou la grille proposée aux élèves (critère 2),
- le mode de communication utilisé dans le questionnaire ou la grille (critère 3),
- la qualité des activités métacognitives (critère 4).

Critère 1 : *objectif(s) pédagogiques de l'outil proposé aux élèves*

(1) Mieux s'approprier un savoir ou un savoir-faire cognitifs bien spécifiques. OUI / NON
(ex : le calcul de la surface du parallélogramme)

(2)	Acquérir ou développer une compétence cognitive globale et transférable *(ex : la synthèse, la prise de notes, la compréhension de textes...)*	OUI / NON
(3)	Développer des comportements socio-affectifs présumés favorables à l'activité cognitive.	OUI / NON
(4)	Développer l'une ou l'autre composante de la métacognition *(ex : l'auto-évaluation)*	OUI / NON
(5)	Apprendre globalement à piloter et à gérer ses apprentissages : anticiper, analyser, auto-évaluer, réguler... *(ex : démarche d'apprentissage d'une leçon, gestion d'une session d'examen...)*	OUI / NON

Critère 2 : *facettes du fonctionnement cognitif et/ou métacognitif concernées par la grille ou le questionnaire proposés aux élèves*

Un même outil pédagogique proposé aux élèves (questionnaire, grille...) est susceptible de couvrir une gamme plus ou moins large d'activités cognitives et/ou métacognitives. C'est pourquoi, il est indispensable d'analyser chacun de ses items en particulier afin de tenter de déterminer les facettes du fonctionnement cognitif et/ou métacognitif auxquelles ces items font appel. Nous recommandons concrètement d'attribuer un numéro à chacun des items de l'outil pédagogique analysé et ensuite de reporter ces numéros en face de chacune des sous-rubriques du critère 2. En outre, lors de l'analyse d'un outil pédagogique, il convient chaque fois de se demander si la réponse fournie par l'élève est :

- non induite (N.I.),
- induite par la question posée (I.Q.),
- induite par les différentes possibilités de réponse qui lui sont proposées (I.R.).

Exemples : « Comment as-tu étudié ta leçon de mathématiques ? » est une question non induite. Par contre la question « Décris le plus précisément possible comment tu as étudié ta leçon de mathématiques ? » est supposée induire une activité d'explicitation de la part du sujet, alors que la question « Compare la manière dont tu as étudié ta leçon de mathématique et celle de français » devrait induire une activité d'analyse (comparaison) chez l'élève.

Modalités d'application du critère 2 :

Indiquer les numéros attribués à chacun des items de l'outil pédagogique analysé de la (ou des) rubrique(s) correspondantes(s).

	N.I.	I.Q.	I.R.
(1) Items portant sur des activités cognitives			
- savoir, savoir-faire spécifiques : - produits			
- processus			
- compétences globales: - produits			
- processus			
(2) Items portant sur des activités métacognitives :			
- Explicitation :			
- Mode descriptif : - fonctionnement personnel			
- variables externes			
- Mode évaluatif : - fonctionnement personnel			
- variables externes			
- Analyse :			
- Mode descriptif : - fonctionnement personnel			
- variables externes			
- Mode évaluatif : - fonctionnement personnel			
- variables externes			
- Régulation(s) :			
(3) Divers *(Items de l'outil pédagogique analysé ne portant pas spécifiquement sur une activité cognitive ou métacognitive.)*			
(4) Autres renseignements :			
- Les items relatifs à l'explicitation sont :			
- contextualisés (référence précise à une situation d'apprentissage),			
- décontextualisés.			
- Les items relatifs à l'explicitation portent sur des situations :			
- passées			
- présentes			
- futures (anticipation)			
- Les items relatifs à l'analyse portent sur des situations			
- passées			
- présentes			
- futures (anticipation)			

Ce critère permet d'identifier et de clarifier les habiletés métacognitives que l'élève est supposé maîtriser pour utiliser correctement cer-

tains outils d'(auto)-évaluation diagnostique qui lui sont proposés. Ces habiletés varient en effet largement en fonction du niveau d'approfondissement attendu du diagnostic : simple description ou analyse de la démarche utilisée, mise en évidence ou non de facteurs plus généraux communs à de multiples tâches, indication ou non de remédiations etc...

Critère 3 : *mode de communication utilisé dans le questionnaire ou la grille*

- Mode descriptif, analytique — OUI / NON
- Mode évaluatif — OUI / NON
- Mode prescriptif — OUI / NON
- Recours au questionnement — OUI / NON
- Questions ou critères :
  - fermés — OUI / NON
  - semi-ouverts — OUI / NON
  - ouverts — OUI / NON

Critère 4 : *qualité des activités métacognitives*

La grille ou le questionnaire proposé à l'élève lui permet-il de caractériser la qualité de ses activités métacognitives ? Voici, à titre indicatif, quelques exemples d'indicateurs de qualité des activités métacognitives :

- Pour l'EXPLICITATION :
  - le degré d'exactitude — OUI / NON
  - le degré de précision — OUI / NON

- Pour l'ANALYSE :
  - le nombre d'éléments de son fonctionnement cognitif mis en relation (dans une situation donnée), — OUI / NON
  - le degré d'élaboration des relations rapportées *(ex : simples relations temporelles, ou relations* but/moyen, condition/effet...) — OUI / NON
  - la validité des explications rapportées — OUI / NON

- Pour la CONCEPTUALISATION :
  - pouvoir décontextualiser une stratégie — OUI / NON
  - pouvoir préciser le domaine d'application d'une stratégie (ne pas lui donner une extension trop faible ou trop grande). — OUI / NON

- Pour l'AUTO-EVALUATION :

  - degré de réalisme (fonder son jugement sur une analyse de la réalité, choisir des indicateurs pertinents, des procédures de vérification, éviter une tendance systématique à la sous ou sur-estimation...) OUI / NON

  - équilibre dans la recherche d'explications de la réussite ou de l'échec entre les facteurs internes et externes. OUI / NON

Dans les faits, ce critère est particulièrement difficile à opérationnaliser.

4. EN GUISE DE CONCLUSION PROVISOIRE

Notre objectif dans cet article a été d'essayer d'articuler les pratiques éducatives qui visent au sens large à aider les élèves à devenir plus autonomes (à «piloter» et à évaluer leurs apprentissages) aux travaux de recherche portant sur la métacognition. Pour ce faire, nous avons tenté dans un premier temps de clarifier le concept de métacognition et d'en proposer un modèle synthétique adapté aux situations éducatives. Celui-ci nous a permis, dans un second temps, d'élaborer un outil d'analyse de ces pratiques éducatives qui soit cohérent par rapport au modèle proposé. Nous espérons ainsi aider les enseignants à réfléchir aux différents types d'activités métacognitives qu'ils tentent de susciter chez les élèves lorsqu'il utilisent ces outils pédagogiques.

Toutefois, nous sommes bien conscient que notre instrument d'analyse de ces pratiques éducatives n'est à ce jour qu'une première ébauche, qu'il conviendrait d'enrichir (en le confrontant à des pratiques éducatives contrastées) et d'opérationaliser davantage. En particulier, dans la suite de nos travaux, nous devrons davantage prendre en compte :

- le contexte pédagogique global dans lequel s'inscrit l'utilisation de l'outil pédagogique utilisé (degré d'appropriation par l'élève ? transfert ?...);

- les facteurs affectifs ou conatifs pouvant interagir avec les aspects cognitifs ou métacognitifs examinés : sentiment d'efficacité personnelle, self-esteem, facteurs attributionnels de la réussite ou de l'échec... Les travaux de Scallon (1992) ouvrent des pistes intéressantes à cet égard.



IMPLICATIONS PSYCHOMÉTRIQUES



Mesure critériée et théorie de la généralisabilité : applications aux mesures cognitives

Dany Laveault

1. LE PROBLÈME DE LA STABILITÉ EN THÉORIE OPÉRATOIRE

La psychologie du développement de l'intelligence de Jean Piaget a principalement contribué à faire ressortir les caractéristiques invariantes de l'intelligence. Inhelder et Piaget (1955) ont bien démontré que malgré la diversité des expériences cognitives individuelles, la structuration des opérations formelles en une organisation adaptée procède par des mécanismes et aboutit à des schèmes qui sont les mêmes pour tous les individus.

À l'opposé de la psychologie génétique, la psychologie différentielle insiste davantage sur les facteurs de l'intelligence qui font qu'un individu peut manifester un meilleur fonctionnement dans certaines activités que dans d'autres. La psychologie différentielle tient compte du fait que la capacité à structurer un problème peut être affectée par des facteurs d'habiletés cognitives tels que la tâche et le contenu.

En théorie opératoire, il est tout aussi nécessaire de s'intéresser, tout comme en psychologie différentielle, aux conditions qui affectent la stabilité de la performance. Des mesures stables sont nécessaires pour plusieurs raisons :

1. d'abord, pour expliquer des écarts de performance qui ne peuvent être imputés à des différences au niveau de la structuration des opérations;
2. ensuite, pour rendre plus pratique et plus réaliste l'utilisation de la théorie opératoire en psychologie et en éducation.

Comment en effet expliquer les différences (p.e. entre individus qui semblent pourtant parvenus aux mêmes paliers de développement) si l'on ignore l'intervalle de confiance à l'intérieur duquel on peut raisonnablement situer les mesures développementales ?

L'étude de la stabilité de la performance est donc d'une grande importance pour expliquer les aspects à la fois réguliers (dans sa structuration) et irréguliers (dans son fonctionnement) du développement. Piaget a porté une attention particulière à l'étude des processus structuraux mais l'aspect fonctionnel, laissé en friche, soulève encore de nombreux problèmes dans l'étude du développement. Voici quelques exemples de ces problèmes :

1. *Le problème des décalages horizontaux* : comment se fait-il qu'une opération développée dans un domaine de la connaissance se généralise plus ou moins facilement à d'autres domaines ?
2. *Le problème de la familiarité avec le contenu* : comment l'expérience accumulée dans un domaine facilite-t-elle ou accélère-t-elle le développement dans ce domaine ?
3. *Le problème de l'hétérogénéité du développement chez un même individu :* comment expliquer qu'un individu ne soit pas parvenu au même niveau opératoire dans différents domaines et jusqu'à quels

«écarts de développement» doit-on s'attendre chez le même individu ?

Les principaux efforts de recherche effectués jusqu'à présent n'ont réussi qu'à décrire les phénomènes. Les tentatives d'explication sont demeurées incertaines pour plusieurs raisons :

- l'absence de devis expérimentaux pouvant contrôler simultanément plusieurs variables affectant la performance;
- le manque de concordance dans la mesure du développement que ce soit au niveau des épreuves employées ou des critères de cotation des résultats;
- le caractère pas toujours approprié de la méthode clinique dans l'étude de la stabilité de la performance.

1.1 Le rôle de la fiabilité en développement

Malgré les difficultés que peuvent occasionner l'instabilité des performances mesurées avec des instruments opératoires, peu d'efforts ont été déployés pour mesurer la fiabilité de ces mesures et des procédures de cotation. Or, avant même de pouvoir discuter de l'hétérogénéité des performances d'un individu ou de l'effet de certaines variables sur sa performance, il faut pouvoir démontrer que les procédures et instruments que nous employons permettent de recueillir une information fidèle. En effet, plusieurs recherches en développement ne fournissent aucune des informations suivantes :

- *Le degré de concordance des évaluateurs :* dans une situation d'interrogation clinique, dans quelle mesure des administrateurs différents parviennent-ils aux mêmes résultats avec le même individu dans des conditions parallèles (Nyiti, 1982) ?
- *Le degré de stabilité temporelle d'une performance :* dans une situation de mesure répétée dans des laps de temps très courts, quelle est la stabilité de la performance observée ?
- *Le degré d'équivalence des situations-problèmes* : dans quelle mesure des problèmes jugés formellement «équivalents» le sont-ils vraiment en situation de mesure répétée ?

Il serait donc important que l'utilisation des épreuves piagétiennes s'accompagne de renseignements permettant d'apprécier la fiabilité des résultats. Ces renseignements devraient inclure :

- *Le coefficient de cohérence interne* (alpha de Cronbach) pour les épreuves collectives ou encore un coefficient de concordance inter-juges pour des interrogations cliniques individuelles. Cette information est nécessaire pour déterminer si items ou juges mesurent tous la même habileté opératoire, et ce, de la même façon.
- *Le coefficient de stabilité,* calculé par la méthode test-retest, pour les études longitudinales. Cette information est essentielle pour départa-

ger le manque de stabilité liée à l'erreur de mesure au test du manque de stabilité liée au développement.

- *Le coefficient d'équivalence,* calculé par la corrélation entre épreuves parallèles : cette information est essentielle pour départager hétérogénéité due à la performance du sujet et hétérogénéité dans l'échantillonnage des items, exercices ou situations-problèmes.
- *Le coefficient de stabilité-équivalence*, calculé par la corrélation entre deux épreuves parallèles administrées à des moments différents : cette information est essentielle pour départager dans la variance des résultats, celle qui est due au développement de celle qui est imputable à la fois à la répétition de la mesure et à l'échantillonnage des items.

On peut se demander si cette tentative d'appliquer les indices les plus courants de fiabilité à l'étude du développement ne revient pas à «psychométriser» Piaget pour la deuxième fois. En effet, Tuddenham (1971) parlait pour la première fois de «psychométriser» Piaget en suggérant l'emploi d'épreuves collectives (plutôt qu'individuelles) et le recours à des procédures standardisées (plutôt que cliniques) d'administration des épreuves opératoires. Tout ceci peut sembler un luxe de procédures, étant donné que la théorie du développement a fort bien su progresser en dépit de ces précautions méthodologiques, et que de fait, plusieurs chercheurs n'y ont guère prêté attention.

Si l'étude du développement doit se restreindre aux grandes périodes marquant les principaux changements marquant l'évolution des opérations de la pensée, alors il y a lieu de croire que l'ensemble des mesures précédentes n'est guère nécessaire. Par contre, si à l'intérieur de ces grandes périodes, l'on souhaite discriminer plus finement l'influence de facteurs fonctionnels d'actualisation de ces opérations, alors il est primordial de réduire l'erreur de mesure. Cette condition est nécessaire dès qu'il est question chez un même individu de discerner son habileté, non plus dans un seul domaine, mais dans plusieurs domaines à la fois, ou encore de différencier, entre individus appartenant à une même période, les facteurs qui facilitent ou entravent le développement. Dans toutes ces situations, l'instrument de mesure doit être fiable, puisqu'une erreur de mesure trop grande risque de voiler des différences réelles entre individus ou encore de masquer, chez un même individu, des différences importantes de réussites dans des domaines différents.

Le développement opératoire et la maturation constituent des effets expérimentaux forts comparativement à des effets plus faibles tels que ceux rapportés dans la littérature et reliés aux effets dûs à la tâche (Linn et Levine, 1976), au contenu (Karplus et Peterson, 1970; Lawson, 1974) et au degré de familiarité du sujet avec le domaine. Ces effets sont plus facilement voilés lorsque, utilisant des instruments et des procédures moins rigoureuses, le chercheur confond les différences individuelles avec l'erreur de mesure aléatoire. Toutefois, ces effets fonctionnels sont particulièrement importants en ensei-

gnement où c'est davantage de micro-développement dont il est question et où l'enseignant ne peut attendre le bon vouloir des processus d'équilibration et de maturation pour faire progresser l'élève. La connaissance de ces effets faibles et du rôle qu'ils jouent dans la consolidation du développement et l'actualisation des opérations dans différents domaines est donc un objet d'étude valable en lui-même.

1.2 Rôle d'une mesure critériée dans l'étude de la stabilité

L'étude du fonctionnement intellectuel requiert non seulement une plus grande rigueur dans l'utilisation des instruments de mesure et des procédures d'évaluation du développement opératoire, mais aussi un plus grand contrôle sur les variables susceptibles d'influencer la performance des sujets. Ceci signifie que l'élaboration et la construction de tests devront porter une attention particulière à la spécification du domaine des items ou des situations-problèmes. Le chercheur en développement cognitif aura donc avantage à utiliser l'une des nombreuses techniques de spécification de domaine à sa disposition en mesure critériée (Roid et Haladyna, 1982).

Les épreuves de type piagétien, individuelles ou collectives, contrôlent un nombre limité de facteurs affectant la situation dans laquelle est réalisée la performance. C'est pourquoi la plupart de ces instruments sont unidimensionnels : c'est-à-dire qu'ils portent sur des items de mêmes univers de contenu et de tâche. Seul le niveau opératoire des items est varié afin d'obtenir une échelle hiérarchique parfaitement reproductible au sens défini par Guttman (1944). Il est alors possible de situer le sujet à l'intérieur d'une hiérarchie développementale qui ne possède cependant qu'une généralisabilité restreinte à d'autres univers de tâches et de contenus.

En comparant le positionnement respectif de plusieurs individus sur un grand nombre d'échelles unidimensionnelles, plusieurs chercheurs ont observé une certaine hétérogénéité des niveaux de performance aux différentes échelles (Stone et Ausubel, 1969; Arlin, 1975; Karplus et al, 1977; Roberge, 1976, Neimark, 1975; Martorano, 1975; Berzonski, 1971). Les épreuves n'ayant pas contrôlé les conditions de réalisation de la performance et l'intervalle de confiance des mesures n'étant pas précisé, il est alors très difficile d'expliquer de façon univoque ces écarts entre résultats à des épreuves unidimensionnelles.

Scallon (1981, 1982) propose un modèle d'évaluation diagnostique fondé sur l'étude des composantes de la variance d'un test construit selon des facettes. Ce modèle permet d'identifier les facettes d'un problème ou d'une situation d'apprentissage qui exercent un effet significatif. Il permet de considérer également certains effets d'interaction. Par exemple, une interaction entre les sujets et les différences obtenues parmi les éléments d'une facette permettent d'inférer qu'une facette ne crée pas les mêmes difficultés chez tous les individus.

Gray (1978) propose d'utiliser conjointement mesure critériée et théorie opératoire. Selon Gray, la théorie de Piaget et la mesure critériée se complètent parce que : «piagetian theory emphasizes the reasoning demanded by an item rather than its content alone». Les épreuves développementales de Piaget permettent de déterminer le type d'opérations employées dans la résolution d'un problème. Pour sa part, la mesure critériée contribue à spécifier le contenu du domaine ou les différentes tâches pour lesquelles ces opérations peuvent être réalisées avec succès. La spécification de domaine des items permet de construire des instruments de mesure opératoire qui contrôlent mieux les facteurs affectant la performance. Le contrôle de ces facteurs est indispensable à une étude scientifique du fonctionnement intellectuel.

Lautrey, Ribeaupierre et Rieben (1981) ont fait l'hypothèse que la variabilité du niveau opératoire d'un même individu en fonction des situations peut renseigner sur des aspects du fonctionnement cognitif que la théorie opératoire ne prend pas en ligne de compte. Toutefois, toutes les manifestations de la variabilité intra-individuelle du niveau opératoire n'ont pas selon eux le même intérêt : elles doivent être suffisamment marquées et stables pour donner lieu à des faits répétables.

L'étude de la variabilité intra-individuelle des réussites passe, selon Lautrey et al. (1981) et selon Londeix (1985), par la réintégration dans la théorie opératoire des aspects négligés de la psychologie différentielle de l'intelligence. En effet, la description des isomorphismes de structure des tâches ne suffit pas à expliquer les variabilités intra-individuelles de performance. Londeix (1985) considère pour sa part que c'est au niveau du choix des situations et du contexte dans lequel les sujets sont placés, ainsi que dans la manière dont leurs réponses sont traitées et cotées que peut s'effectuer le mieux cette interconnection entre psychologie opératoire et psychologie différentielle.

2. VERS UN MODÈLE HEURISTIQUE DE MESURE DE LA STABILITÉ

2.1 Difficultés soulevées par la théorie classique

L'applicaton de la mesure critériée à l'étude du développement pose le problème de l'estimation de la stabilité de la performance, d'un ou plusieurs individus, dans plusieurs contextes de tâches, de contenus et de niveaux opératoires d'organisation. Selon Popham et Husek (1979), les indices de fiabilité fondés sur la cohérence interne ne sont d'aucune utilité à la mesure critériée. Ils prétendent que puisque la mesure critériée a pour but de déterminer le rendement d'un individu par rapport à un critère, la signification du score de l'individu ne devrait pas dépendre des scores d'autres individus. Ils en concluent que la variabilité des scores n'est pas une condition nécessaire à la création d'un bon test critérié.

Les indices classiques de cohérence interne, stabilité, équivalence et de stabilité-équivalence ne permettent pas non plus d'aborder la question de la stabilité de la performance lorsque plus d'une variable indépendante est en jeu. Chaque effet est évalué séparemment sans possibilité d'étudier les interactions entre variables. Pris individuellement, ces indices nous renseignent sur l'erreur de mesure liée à la répétition dans le temps (coefficient de stabilité), à l'échantillonnage des items (coefficient d'équivalence) ou à l'homogénéité des items (cohérence interne). Cependant, ils ne permettent pas de tenir compte simultanément des conditions affectant la stabilité de la performance ni même de leurs interactions possibles.

L'utilisation d'indices multiples de fiabilité ne permet pas d'apporter de solutions valables à un problème relativement simple et fort susceptible de se produire en recherche. Il est donc essentiel de pouvoir situer les sources de variation dans un cadre unique pour apprécier simultanément l'importance de plusieurs facettes d'un instrument de mesure ainsi que leurs interactions.

2.2 La théorie de la généralisabilité

Cronbach, Rajaratnam et Gleser (1963) ont élaboré la théorie de la généralisabilité dans le but de réunir en un seul concept les différentes définitions de la fiabilité. En utilisant les principes de l'analyse de variance, Cronbach et al. proposent de quantifier l'importance de chaque source de variation d'une situation de mesure. Le score vrai est l'espérance mathématique de toutes les observations possibles et l'erreur est le résultat d'une fluctuation dans l'échantillonnage de certains niveaux de facettes considérés (évaluateurs, moments, formes d'items, etc.).

La généralisabilité est donc un concept plus englobant que celui de fiabilité. Il permet également de décrire des situations de mesure plus complexes et plus près de la réalité. Cardinet et Tourneur (1985) la définissent ainsi : «La généralisabilité est donc le degré auquel on peut généraliser un résultat obtenu dans des conditions particulières à la valeur théorique recherchée».

Il existe donc autant de coefficients de généralisabilité qu'il existe de plans d'observation (ou plans d'expérience). Néanmoins, ceux-ci ne sont pas tous d'un égal intérêt. Selon l'univers de généralisation choisi, l'estimation sera plus précise selon que cet univers est étroit ou étendu. Cependant, il y a peu d'utilité à généraliser une information à un univers restreint de conditions.

En se fondant sur le principe de la symétrie des données dans un plan d'analyse de variance, Cardinet et Tourneur (1985) ont étendu la théorie de la généralisabilité initiale (Cronbach, Gleser, Nanda et Rajarathnam, 1972). En effet, pour Cronbach et al. la facette «sujets» constitue la seule facette de différenciation utile. Or, en psychologie et en éducation, le chercheur n'est pas seulement intéressé à estimer la stabilité des scores des sujets, il peut être intéressé par d'autres facettes de différenciation tels que les items. Il peut

s'agir d'estimer la stabilité des effets des différentes tâches et des différents contenus des items introduits dans le plan d'observation. Dans de telles conditions, ce ne sont plus les sujets que l'analyse de généralisabilité cherchera à différencier, mais bien les tâches et les contenus en tant qu'objets d'observation. L'échantillonnage des personnes est alors considéré comme erreur de mesure.

La démarche proposée par Cardinet et Tourneur (1985) permet donc de définir une série de procédures de calcul applicables à tous les types de plans expérimentaux et qui permettent de tenir compte du «projet de mesure». En effet, selon ces deux auteurs :

> *L'erreur n'apparaît que par rapport à un projet de mesure. Elle suppose une intention particulière qui privilégie une ou plusieurs facettes comme conditions d'observation, c'est-à-dire comme sources d'erreurs... C'est (...) après le choix d'une direction privilégiée de mesure, que s'insère la théorie de la généralisabilité. Son rôle est de préciser l'importance de la variance due aux facettes privilégiées (variance de différenciation) par rapport à la variance due à l'échantillonnage des conditions d'observation (variance d'erreur).* (p.31)

La démarche proposée par Cardinet et Tourneur (1985) s'effectue en quatre phases : les phases 1 et 2 se rapportent à l'analyse de variance; les phases 3 et 4 se rapportent à l'étude de généralisabilité proprement dite. Voici une courte description de ces quatre phases :

1. *Plan d'observation :* au cours de cette phase, on procède au choix des facettes et au nombre de niveaux de chaque facette. On précise également les interrelations (nichage, croisement) entre ces facettes.
2. *Plan d'estimation* : au cours de cette phase, on détermine quelles facettes représentent un échantillonnage aléatoire ou exhaustif (effet fixe) de niveaux.
3. *Plan de mesure* : au cours de cette phase, on identifie quelles facettes sont liées au projet de mesure (facettes de différenciation) et quelles facettes seront considérées comme sources d'erreur de mesure (facettes d'instrumentation). C'est au cours de cette phase que les composantes de variance calculées à la phase deux peuvent être attribuées à la variance vraie ou à la variance d'erreur, permettant ainsi le calcul du coefficient de généralisabilité.
4. *Plan d'optimisation* : cette phase consiste à modifier soit le plan d'observation, soit le plan d'estimation, soit le plan de mesure ou encore une combinaison des trois afin de maximiser la généralisabilité des observations. Le chercheur devra trouver alors un équilibre entre précision de la mesure et étendue de l'univers de généralisation. En effet, plus l'univers de généralisation est grand, plus il est difficile d'obtenir des mesures peu entachées d'erreur. Par contre, il y a peu d'intérêt pratique à utiliser des mesures précises lorsque l'univers de généralisation est trop étroit.

L'étude de la généralisabilité est utile à un autre point de vue. Elle permet de tenir compte du type d'erreur de mesure dans le calcul du coefficient de fiabilité. Cardinet et Tourneur (1985) établissent une distinction entre erreur de mesure absolue et erreur de mesure relative. Cette distinction est primordiale en psychologie du développement ainsi que dans tout système de mesure psychologique où des seuils critiques sont utilisés pour déterminer si un stade ou une étape de développement a été atteint ou pas. Dans de tels cas, il est nécessaire de calculer la fiabilité par rapport à l'erreur absolue de mesure et non pas par rapport à l'erreur relative.

En psychologie différentielle, c'est l'erreur sur les positions relatives des résultats qui est la plus pertinente. En psychologie du développement, il n'est pas utile de déterminer si un instrument permet de mieux distinguer si un individu est plus avancé dans son développement qu'un autre individu, car ce genre de tests n'est pas utilisé à des fins de classement ou de sélection comme c'est le cas par exemple de tests d'intelligence de type Q.I. Dans le contexte développemental, c'est davantage l'erreur sur la valeur obtenue par rapport à une échelle dont les échelons sont définis a priori qui doit intéresser le chercheur. Afin de déterminer si la performance d'un individu a dépassé une valeur critique permettant d'affirmer qu'un nouveau stade ou étape de développement a été atteint, le chercheur a besoin d'une estimation du degré de réussite à l'intérieur d'un intervalle de confiance le plus restreinte possible.

2.3 Objectifs

Afin d'illustrer comment il est possible d'appliquer la théorie de la généralisabilité au calcul de la fiabilité dans une variété de conditions, nous avons choisi un test construit au moyen de facettes. Nous démontrerons comment l'étude des composantes de variance et l'optimisation des différentes mesures permettent d'approfondir le diagnostique des difficultés.

Les objectifs de cette étude sont les suivants :

1. Illustrer comment l'évaluation diagnostique peut tirer profit de l'étude des différentes composantes de variance des résultats au test, pour mieux faire porter l'intervention et la remédiation sur les facteurs les plus utiles.
2. À partir du principe de symétrie des données de l'analyse de variance, étudier la fiabilité d'un instrument selon différents projets de mesure. Dans certaines situations, il faut considérer l'erreur de mesure sur les facettes d'intérêt (spécification du domaine des items) plutôt que sur les sujets. Les résultats de l'étude de généralisabilité seront donc présentés quant à plusieurs plans de mesure.
3. Faire ressortir l'importance d'utiliser l'erreur de mesure absolue dans l'étude du développement. Un instrument peut être adéquat à des fins de classement, mais avoir peu de fiabilité lorsqu'il s'agit de déterminer un seuil de réussite. C'est pourquoi des coefficients de généra-

lisabilité d'erreur relative et d'erreur absolue seront calculés pour différents plans de mesure.

4. Comparer l'efficacité de différents plans d'optimisation. Il s'agira de démontrer comment l'étude de généralisabilité permet au chercheur de mieux optimiser sa mesure en sachant quels types d'items et quelle quantité d'items additionnels permettent d'accroître la fiabilité du test en entier.

3. MÉTHODOLOGIE

3.1 Sujets

Quarante-huit sujets adultes, des niveaux primaire (8), secondaire (27) et collégial (13) ont répondu au test Imageo5D sur une base volontaire. Ces adultes, 31 femmes et 17 hommes étaient âgés entre 19 et 55 ans, la moyenne d'âge étant de 29,5 ans (s = 6,96).

3.2 Instrument

L'instrument Imageo5D (Laveault, 1987) est un instrument critérié de mesure de l'habileté à représenter mentalement différents arrangements de figures géométriques dans un espace plan. À partir d'un ensemble de 5 ou de 7 figures géométriques dont une partie ou la totalité sont voilées, le sujet doit reconstituer les figures individuelles afin de déterminer quels ensembles sont ou ne sont pas constitués des mêmes figures.

La tâche de ce test consiste à trouver deux ensembles constitués des mêmes figures de base. Ces figures de base sont celles du casse-tête chinois «Tangram» : 2 petits triangles, 1 triangle moyen, 1 parallélogramme, 1 carré pour l'ensemble de 5 figures et 2 grands triangles en plus, pour l'ensemble de 7 figures. Ces figures peuvent être combinées de plusieurs façons pour reproduire une infinité de formes différentes. Le test introduit la difficulté suivante : l'aire de chaque figure est proportionnelle à l'aire de chacune des autres figures et chaque figure plus grande peut être reconstruite à partir d'une combinaison de figures plus petites. C'est ainsi que deux petits triangles peuvent être combinés pour former un carré, un parallélogramme et un triangle moyen, etc.

Sur le plan opératoire, ce test nécessite des habiletés de représentation spatiale dans l'espace plan ainsi que l'opération d'arrangement. Chaque figure plus grande peut-être décomposée pour constituer des figures plus petites et réciproquement, deux figures plus petites ou plus peuvent reconstituer une figure plus grande. Le sujet doit déterminer, à partir de l'ensemble des figures en présence (non voilées) quels arrangements de ces figures sont possibles dans un plan.

Ce test requiert les opérations concrètes lorsqu'il s'agit d'effectuer des arrangements d'un petit nombre de figures. Par contre, les opérations formelles deviennent nécessaires, lorsqu'en présence d'un grand nombre de figures voilées, le sujet doit envisager un ensemble d'arrangements possibles et éliminer ceux qui ne peuvent être construits à partir de l'ensemble des figures géométriques de base (cinq ou sept).

Ce test permet de contrôler trois facettes de résolution de problèmes. Ces facettes forment un plan d'observation totalement croisé et sont décrites dans la figure 1. Ces trois facettes sont les suivantes :

- Contours :
  1. détaillés;
  2. non détaillés.
- Nombre de figures :
  1. sept;
  2. cinq.
- Nombre d'indices :
  1. trois figures indices;
  2. deux figures indices;
  3. aucune figure indice.

Imageo5D	Exercices avec 7 figures			Exercices avec 5 figures		
	3 figures/indices	2 figures/indices	0 figures/indices	3 figures/indices	2 figures/indices	0 figures/indices
Fgures à contours détaillés						
Fgures à contours non détaillés						

FIGURE 1

Spécification du domaine du test Imageo5D (Laveault, 1987)

Ces facettes permettent de générer 12 catégories d'items différents, chacune répondant à des conditions de réussite à la fois semblables et à la fois

différentes de celles des autres items. Étant donné qu'il s'agit d'un plan parfaitement croisé, il est également possible d'envisager l'étude d'interactions entre les trois facettes et leurs différents niveaux. Un exemple d'item correspondant à la catégorie «contours détaillés» x «sept figures» x «deux figures indices» est illustré à la figure 2.

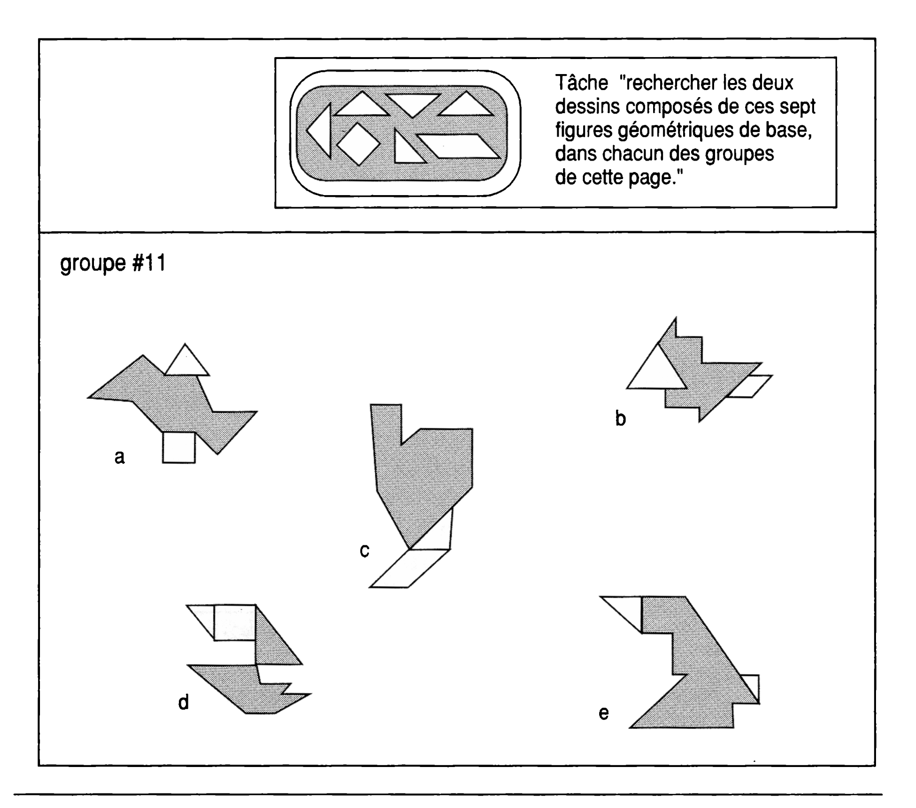


FIGURE 2

Exemple d'item du test Imageo5D (Laveault, 1987)

3.3 Procédure

Le test Imageo5D a été administré dans le cadre d'une recherche plus vaste sur l'évaluation formative des habiletés intellectuelles de l'adulte. Ce test a été le premier d'une série à être administré aux sujets.

Il a été administré durant les périodes de cours à l'institution scolaire fréquentée par le sujet. Chaque sujet recevait 4 $ pour chaque heure de participation à la recherche.

4. RÉSULTATS

4.1 La théorie de la généralisabilité

L'étude de la généralisabilité des résultats du test Imageo5D se fera selon les quatre étapes proposées par Cardinet et Tourneur (1985) : plan d'observation, plan d'estimation, plan de mesure et finalement, plan d'optimisation.

4.2 Plan d'observation et plan d'estimation

Avant d'entreprendre le calcul des composantes de variance et des différents coefficients de généralisabilité pour chacun des objets de mesure, il est nécessaire de spécifier l'univers de généralisation du plan d'estimation. Chaque facette a été estimée de la manière suivante :

1. la facette «contours» (C) a été considérée comme fixée, puisqu'elle épuise toutes les possibilités de niveaux de l'univers de généralisation (contours détaillés et non détaillés);
2. les facettes «nombre de figures» (N), «nombre d'indices» (I) et «personnes» (P) ont été considérées comme aléatoires. Notre objectif consiste à généraliser les conclusions à toutes les situations différentes de «nombre de figures», «nombre d'indices» et «personnes» et non seulement aux conditions rencontrées dans la recherche. Une ou plusieurs de ces trois facettes seront considérées à tour de rôle comme facette(s) d'instrumentation.

Étant donné que la facette «contours» est fixée, le plan d'estimation de cette recherche est totalement croisé avec des effets mixtes.

4.3 Composantes de variance et plans de mesure

Le tableau 1 indique quel pourcentage de la variance totale est expliqué par les composantes de variance de chaque facette dans un plan mixte (C est fixé, les autres facettes sont aléatoires). On y remarque que les composantes de variance les plus importantes sont d'abord la triple interaction PNI et l'interaction PCNI confondue avec l'erreur. Les interactions PI, PNI et PCN sont également des effets non négligeables. Les facettes simples les plus importantes sont d'abord les sujets suivies par le nombre de figures indices et le nombre de figures totales. La facette C est celle qui possède la plus faible composante de variance.

Une importante interaction PI indique que la facette des figures indices n'affecte pas la performance de chaque sujet de la même manière. Les triples interaction PNI et PCN indiquent que la performance de chaque sujet, pour les items de 5 ou 7 figures, est affectée par le nombre de figures indices et, dans une moindre mesure, par le détail des contours. L'ensemble de ces in-

teractions s'explique facilement : un sujet habile aura la même performance, peu importe le nombre de figures totales ou le nombre de figures indices. Il en sera de même pour un sujet éprouvant de sérieuses difficultés. Par contre, pour un sujet d'habileté «intermédiaire», le nombre de figures indices peut n'être d'aucune utilité lorsque le nombre total de figures est restreint (5 figures) mais peut, par contre, devenir fort utile lorsque le problème comporte un nombre plus grand de figures.

TABLEAU 1

Composantes de variance du test Imageo5D[1]

Source de variation	Somme des carrés	dl	Carré moyen	Espérance de variance modèle mixte	Composante de variance
P	38,826	47	0,826	0,0523	18,77
C	1,000	1	1,000	0,0030	1,07
PC	6,667	47	0,142	-0,0012*	0,00
N	4,000	1	4,000	0,0135	4,83
PN	7,333	47	0,156	-0,0013*	0,00
CN	0,174	1	0,174	-0,0029	0,00
PCN	7,493	47	0,159	0,0064	2,29
I	16,253	2	8,127	0,0414	14,87
PI	19,413	94	0,207	0,0107	3,83
CI	0,198	2	0,099	-0,0045*	0,00
PCI	11,135	94	0,118	-0,0007*	0,00
NI	0,260	2	0,130	-0,0004*	0,00
PNI	15,406	94	0,164	0,0819	29,42
CNI	1,941	2	0,970	0,0088	3,18
PCNI	11,392	94	0,121	0,0606	21,75

* Les espérances de variance négatives sont ramenées à 0 tel que recommandé par Cardinet et Tourneur (1985), page 89.

L'étude des composantes de variance fournit également des informations diagnostiques précieuses pour la mise au point d'un plan d'intervention. La présence d'une forte composante d'interaction PNI (29,42%) indique que l'intervention devra être individualisée. En effet, si l'effet des composantes de variance de la facette «nombre de figures indices» (14,87%) indique que le

1 Calculs effectués au moyen du programme ETUDGEN (Dufour, 1987).

fait d'augmenter le nombre de figures indices facilite la tâche pour l'ensemble des élèves, la triple interaction PNI révèle qu'elle dépend encore plus de l'élève en cause et du nombre de figures (5 ou 7) sur lesquelles il travaille. Pour un élève éprouvant des difficultés, il faudrait donc considérer quels items comportant 5 ou 7 figures sont réussis avant de décider si l'on réduit le nombre de figures indices, le nombre total de figures ou les deux à la fois. La présence d'une forte interaction PNI tend à indiquer que l'on ne saurait utiliser le même plan d'intervention pour faire progresser simultanément tous les sujets. Ce qui s'avère une aide pour l'un, risque de ne pas l'être pour l'autre.

L'étude des composantes de variance associées à la facette «contours» révèle que celle-ci n'influence que très peu les résultats des sujets de notre échantillon (1,07%). Il y a donc peu d'intérêt à manipuler cette variable : augmenter ou diminuer le nombre de contours des figures cachées ne semble pas aider les sujets de notre échantillon outre mesure. On peut donc l'ignorer dans notre plan d'intervention avec des sujets de ce groupe.

Enfin, l'ordre de grandeur de la composante «personnes» (18,77%) nous indique l'homogénéité du groupe. Une composante élevée indique des différences importantes entre les scores des sujets. Une composante faible révèle que les sujets réalisent des performances semblables pour les questions tirées de l'univers de généralisation du test. Dans le cas qui nous intéresse, le groupe est relativement hétérogène. Pour évaluer l'efficacité de notre plan d'intervention, nous pourrions réexaminer cette composante de variance suite à une intervention pédagogique. Une diminution du pourcentage de variance associé à cette facette indiquerait que le groupe devient plus homogène et donc, que nous avons réussi à réduire les écarts de performance entre ceux qui réussissaient bien et ceux qui éprouvaient auparavant des difficultés.

4.4 Coefficients de généralisabilité

Le tableau 2 présente un exemple de calcul du coefficient de généralisabilité des scores des sujets pour le plan P/CNI. Si le but recherché est de classer les sujets par ordre d'importance de leur score total, comme c'est le cas lors de tests normatifs ou de tests de sélection, alors le coefficient de généralisabilité d'erreur relative est passablement acceptable (0,7520) considérant le petit nombre d'items (12) ayant servi à la mesure. Par contre, si l'on devait effectuer, dans le contexte d'une mesure critériée, une prédiction quant à la proportion des items de l'univers de généralisation que le sujet est capable de réussir, alors le coefficient de généralisabilité calculé sur l'erreur absolue de mesure est beaucoup moins acceptable (0,5810). Or, dans le cas d'une mesure développementale, telle que celle du niveau opératoire, il importe peu de déterminer quel sujet est le plus avancé dans son développement. Il importe davantage de déterminer avec quel degré de précision il est possible d'estimer si le sujet a dépassé un seuil absolu de réussite correspondant à un stade précis de développement. Dans cette situation, il faut tenir compte de l'erreur absolue dans le calcul du coefficient de généralisabilité. Avec 12 items,

les résultats au test Imageo5D ne peuvent pas être considérés comme suffisamment fidèles pour déterminer avec assurance le stade de développement auquel est parvenu chaque sujet.

Si l'on considère que bien des épreuves piagétiennes ne comportent guère plus de 12 items, on peut s'interroger sur la possibilité de classer correctement en stades un individu à partir d'un si petit échantillon d'items. Pour cela, il faudrait effectuer des calculs similaires à celui-ci pour des épreuves piagétiennes connues afin de vérifier si l'erreur absolue de mesure est du même ordre de grandeur avec le même nombre réduit d'items.

TABLEAU 2

Exemple de calcul de la généralisabilité : plan P/CNI[2]

Facettes de différenciation		Facettes de différenciation		
Source	Variance	Source	Erreur relative	Erreur relative
P	0,0523	C		0,0000
		PC	0,0000	0,0000
		N		0,0067
		PN	0,0000	0,0000
		CN		0,0000
		PCN	0,0000	0,0000
		I		0,0138
		PI	0,0036	0,0036
		CI		0,0000
		PCI	0,0000	0,0000
		NI		0,0000
		PNI	0,0137	0,0137
		CNI		0,0000
		PCNI	0,0000	0,0000
		C		0,0000
Variance totale	0,0523		0,0172	0,0378
Coefficient de généralisabilité[a]			0,752	0,581

a. Le coefficient de généralisabilité est le rapport entre la variance de différenciation (variance attendue des scores univers) et la somme des variances de différenciation et d'erreur (variance attendue des scores observés).

2 Calculs effectués au moyen du programme ETUDGEN (Dufour, 1987)

Le tableau 3 indique les coefficients de généralisabilité pour plusieurs plans de mesure. Ces coefficients sont calculés soit à partir des composantes de variance d'erreur absolue, soit à partir des composantes de variance d'erreur relative.

Ce tableau permet d'illustrer la grande flexibilité de la théorie de la généralisabilité. En effet, alors qu'en théorie classique, le score du sujet est le seul objet de mesure considéré, plusieurs objets de mesure peuvent être envisagés dans le modèle de la théorie de la généralisabilité proposé par Cardinet et Tourneur (1985). En effet, le chercheur peut être intéressé non seulement par l'erreur de mesure sur le score des sujets, mais aussi par l'erreur de mesure sur le degré de difficulté de certains éléments des facettes qu'il a introduits dans une résolution de problème.

TABLEAU 3

Valeurs de généralisabilité par plusieurs plans de mesure[3]

(1) Plans de mesure	(2) $\rho\delta$	(3) $\rho\Delta$
(P/CNI)	0,752	0,581
(I/CNP)	0,974	0,832
(N/CIP)	0,959	0,464
(C/NIP)	0,629	0,113
(PI/CN)	0,718	0,686
(PN/CI)	0,681	0,595
(NI/CP)	0,966	0,948

Légende : (1) (différenciation/instrumentation)
(2) coefficient de généralisabilité (erreur relative)
(3) coefficient de généralisabilité (erreur absolue)

Dans le cas du test Imageo5D, on peut se demander, par exemple, quelle est la fiabilité des niveaux de la facette I ? Jusqu'à quel point peut-on être assuré que la difficulté relative (l'ordre de difficulté) ou la difficulté absolue (le taux de réussite et d'échec dans la population) de problèmes comprenant un nombre variable de figures indices peuvent être mesurées avec fiabilité par le test ? Ce type de plan de mesure requierre que nous considérions la facette I comme facette de différenciation (facette mesurée) et les autres facettes (C, N et P) comme facettes d'instrumentation décrivant l'univers de généralisation. Il cherche à répondre à la question suivante : quelle est la fiabilité du niveau de difficulté des éléments de la facette I, peu importe le

3 Calculs effectués au moyen du programme ETUDGEN (Dufour, 1987)

nombre de figures totales (N), le type de contours (C) ou encore le type de sujets (P) ?

Le tableau 3 permet de répondre à la question précédente, de même qu'à toutes les questions similaires qui utiliseraient une autre facette ou l'interaction entre deux facettes comme objet de mesure. Dans ce tableau, les plans de mesure sont décrits selon la notation employée par Cardinet et Tourneur (1985). À gauche de la barre oblique, figurent les facettes de différenciation (ou de mesure). À droite de la barre oblique, figurent les facettes d'instrumentation (ou de généralisation). En ce qui concerne la fiabilité de la difficulté des éléments de la facette I (plan I/CNP), les résultats indiquent que la fiabilité d'erreur relative (0,974) et la fiabilité d'erreur absolue (0,832) sont fort acceptables. Il n'en est pas de même pour toutes les facettes, selon le plan de mesure considéré.

Les résultats du tableau 3 indiquent que l'erreur relative est moindre lorsque le plan de mesure considère le nombre total de figures indices (0,9737), le nombre total de figures (0,9594) ou encore l'interaction entre ces deux facettes (0,9620) comme objet de mesure ou si l'on préfère, comme variance de différenciation. Le classement du degré de difficulté des items en fonction du nombre de figures indices, du nombre total de figures ou encore de l'interaction entre ces deux facettes est relativement stable. Cependant, si le but visé est d'estimer la valeur absolue de la difficulté des items de l'univers de généralisation, alors seuls les coefficients de généralisabilité des plans de mesure (I/CNP=0,8232) et (NI/CP=0,9400) sont acceptables. Le plan de mesure (N/CIP), même s'il permet de classer les items selon leur degré de difficulté avec assurance (0,9594), ne permet pas d'effectuer une mesure précise de la valeur absolue de la difficulté de chaque niveau de la facette «nombre total de figures» (0,4642).

Le calcul des coefficients de généralisabilité du tableau 3 indique également que le degré de difficulté de chacun des niveaux de la facette «contours» est entaché d'une forte erreur relative (0,6293) de même que d'une forte erreur absolue (0,1113). Il semble que cette facette ne représente pas une composante suffisamment grande de la variance totale pour donner lieu à des effets stables qui pourraient intéresser le chercheur. Cette facette devrait être reformulée ou éliminée, pour être remplacée par un plus grand nombre de conditions d'observation d'une autre facette.

Le tableau 3 démontre que l'instrument de mesure Imageo5D convient davantage à certaines situations diagnostiques qu'à d'autres. Les coefficients de généralisabilité d'erreur relative indiquent clairement qu'il ne faut pas beaucoup de personnes, ni de nombreux niveaux des autres facettes du test pour démontrer que plus le nombre total de figures est élevé (N/CIP = 0,959) ou encore plus le nombre total de figures indices est faible (I/CNP = 0,974), plus le problème augmente en difficulté. Par contre, si l'on cherche à calculer le taux de difficulté dans chacune des situations précédentes, la situation est toute autre. Le coefficient de généralisabilité d'erreur absolue indique

qu'il est beaucoup plus facile d'estimer le taux de réussite pour les trois niveaux de figures indices (I/CNP = 0,832) que pour les deux niveaux portant sur le nombre total de figures (N/CIP = 0,464). Enfin, il faudrait accroître de façon considérable l'échantillon de personnes ou les niveaux des autres facettes d'observation pour estimer de façon précise le taux de réussite dû aux contours (C/NIP = 0,113).

Ceci ne sont que quelques exemples des possibilités diagnostiques qui s'offrent au chercheur. Celui-ci peut ainsi déterminer les conditions dans lesquelles il lui sera possible d'estimer de façon fiable, non seulement les scores des sujets, mais aussi le taux de difficulté pour les différentes facettes de résolution de problèmes pour des sujets provenant de la même population. Plus l'effet expérimental est fort, moins il est nécessaire d'avoir un grand échantillon des niveaux de facettes d'instrumentation. Par contre, lorsque l'effet expérimental est faible, le chercheur peut estimer à l'avance la grandeur de l'échantillon nécessaire de l'univers de généralisation pour estimer de façon précise le taux de difficulté des différentes facettes. Il n'est donc pas possible de parler de généralisabilité d'un instrument sans tenir compte du projet de mesure. Dans la pratique, le calcul des coefficients de généralisabilité pour un aussi grand nombre de plans de mesure n'est pas chose courante, ni même souhaitable. Seuls les coefficients de généralisabilité correspondant à des objets de mesure bien précis ont besoin d'être calculés. En effet, le chercheur cherchera avant tout à optimiser les plans de mesure pour lesquels il a un intérêt particulier.

5. DISCUSSION

Les résultats de l'étude de généralisabilité ont permis une analyse détaillée des facettes ayant servi à construire le test Imageo5D. Ils indiquent que les facettes « nombre total de figures » et « nombre de figures indices » constituent des aspects de la résolution de problème qui sont suffisamment stables pour être d'intérêt diagnostique. Par contre, ils indiquent que la facette « contours » ne permet pas d'en arriver à des conclusions généralisables quant à ses effets sur la performance des sujets, quel que soit le nombre total de figures en jeu ou le nombre de figures indices.

Les résultats indiquent également que la mesure relative de la performance des sujets est généralisable, mais pas la mesure absolue. Ceci signifie qu'avec 12 items seulement, il n'est pas possible d'estimer avec une très grande précision le niveau de développement des sujets, même s'il est possible d'identifier avec un degré d'assurance plus élevé quels sont les sujets les plus avancés dans leur développement. Afin de réduire l'erreur absolue de mesure sur le score de chaque sujet, il faut donc accroître l'échantillonnage des items. Toutefois, connaissant l'importance de chaque facette en termes de composante de variance totale, il est possible de distribuer cet accroissement du nombre d'items selon l'importance de chaque facette de l'instrument de me-

sure. Un examen du tableau 4 permettra d'apporter une solution plus adaptée à ce problème.

La méthode la plus simple pour optimiser la mesure consiste à accroître le nombre de niveaux pour les facettes d'instrumentation dont les composantes de variance sont les plus grandes. Pour chaque plan de mesure du tableau 2 (sauf le plan C/INP), le tableau 4 présente les coefficients de généralisabilité que l'on obtiendrait si l'on doublait le nombre des niveaux des facettes d'instrumentation. Cette procédure est en fait l'extension, par la théorie de la généralisabilité, de la formule de Spearman-Brown utilisée en théorie classique des tests.

Les différents plans d'optimisation envisagés permettent de constater que, pour accroître la généralisabilité des scores absolus des sujets, il est préférable d'augmenter l'échantillonnage des questions sur la facette I plutôt que sur la facette N. Si l'on double le nombre de niveaux de la facette I, le coefficient de généralisabilité (erreur absolue) passe de 0,581 à 0,702. Si l'on double le nombre de niveaux de la facette N, le même coefficient passe de 0,581 à 0,655. Toute chose étant égale par ailleurs, il est donc préférable d'augmenter d'abord le nombre de niveaux de la facette I. L'augmentation du nombre de niveaux sur les deux facettes à la fois fournit un coefficient de généralisabilité (erreur absolue) encore plus acceptable (0,772). Cependant, le nombre d'observations totales est passé de 576 à 2304 et le nombre d'items de 12 à 48.

Tous ces plans d'optimisation assument que la facette C serait conservée telle quelle. Or, il serait possible de conserver un coefficient de généralisabilité fort acceptable en éliminant cette facette, ce qui réduirait de 48 à 24 le nombre d'items dans de telles conditions. L'estimation du coefficient de généralisabilité pour un test de 24 items (I = 6; N = 4) est différente selon que l'on utilise les composantes de variance calculées à partir du groupe d'items à contours détaillés (CD) ou encore à partir du groupe d'items à contours non détaillés (CND). Dans le premier cas (CD), les coefficients de généralisabilité seraient de 0,830 (erreur relative) et de 0,692 (erreur absolue). Dans le second cas (CND), les estimations fournissent des résultats légèrement supérieurs : 0,864 (erreur relative) et 0,774 (erreur absolue). Il y aurait donc un certain avantage à employer des items comportant uniquement des ensembles de figures à contours non détaillés.

Le tableau 4 indique également comment en doublant le nombre de niveaux des autres facettes d'instrumentation, il est possible d'accroître la généralisabilité des scores. Les changements les plus importants se produisent lorsque le nombre de niveaux des facettes I et N est doublé. Le tableau 5 ne rapporte pas de résultats pour la facette C : celle-ci ayant été fixée à deux niveaux, il n'est pas possible de faire d'estimation au-delà de cette valeur.

TABLEAU 4

Généralisabilité pour différents plans d'optimisation [4]

Plan de mesure	Optimisation sur (1)						
(2)	**O**	**I**	**N**	**P**	**N**	**IP**	**NP**
N =	(576)	(1152)	(1152)	(1200)	(2304)	(2400)	(2400)
		(3)					
(P/CNI)	0,752	0,859	0,834	____	0,910	____	____
	0,581	0,702	0,655	____	0,772	____	____
(I/CNP)	0,974	____	0,985	0,988	____	____	0,993
	0,823	____	0,890	0,842	____	____	0,908
(N/CIP)	0,959	0,979	____	0,980	____	0.990	____
	0,464	0,618	____	0,479	____	0,640	____
(PI/CN)	0,718	____	0,836	____	____	____	____
	0,686	____	0,814	____	____	____	____
(PN/CI)	0,681	0,810	____	____	____	____	____
	0,595	0,746	____	____	____	____	____
(NI/CP)	0,966	____	____	0,983	____	____	____
	0,948	____	____	0,974	____	____	____

Légende :
(1) facettes dont le nombre de niveaux est doublé
(2) nombre total d'observation
(3) valeur d'erreur relative et absolue
O aucun changement
F facette # indices
N facette # figures
P facette # personnes
C facette # contours

4 Calculs effectués au moyen du programme ETUDGEN (Dufour, 1987)

Conclusion

La théorie de la généralisabilité est une méthode statistique d'étude de la fiabilité mieux adaptée à la situation de la mesure critériée et de la mesure en développement que les indices classiques d'estimation de la fiabilité. Elle permet de tenir compte d'un plus grand nombre de facettes qui affectent la performance des sujets et de par le principe de la symétrie des données de l'analyse de variance, elle permet d'estimer la fiabilité non seulement des scores des sujets mais aussi des niveaux de difficulté des éléments de ces facettes. Lors de l'optimisation de l'instrument de mesure, elle tient compte des facettes qui représentent les composantes de variance les plus importantes en fonction du projet de mesure envisagé. Dans de telles conditions, l'optimisation de la mesure peut s'effectuer selon un rationel plus élaboré. Au lieu d'une démarche «essai-erreur» visant à créer plus d'items que nécessaires et consistant à éliminer par la suite les «mauvais» items, l'étude de la généralisabilité permet, à partir d'un échantillonnage d'items regroupés en facettes, de déterminer les propriétés importantes de ces items dans le projet de mesure.

La théorie de la généralisabilité constitue un outil méthodologique important pour l'étude des aspects fonctionnels du développement. Pour mieux comprendre le fonctionnement des opérations, et incidemment leur développement, il faut être en mesure d'étudier celles-ci dans des conditions variées et stables d'observation. L'étude de généralisabilité permet de déterminer dans quelles conditions la performance d'un individu est généralisable, ou encore de spécifier si des effets de tâche et de contenu sont stables, peu importe les autres conditions (niveaux opératoires, conditions d'entrevue, etc.).

La théorie de la généralisabilité offre beaucoup plus qu'une simple assurance dans la mesure effectuée par le chercheur en développement. C'est une méthodologie de la mesure qui permet un diagnostic précis des conditions affectant la performance du sujet. Elle permet aussi d'améliorer les instruments de mesure cognitive et de mieux identifier les variables affectant le fonctionnement des opérations intellectuelles. Parmi les principaux avantages à retirer de l'étude de la généralisabilité des mesures cognitives, relevons :

1. L'étude des composantes de variance d'interaction et des composantes principales renseigne le praticien et le chercheur sur l'utilité d'une intervention individuelle ou collective. Elle permet également d'identifier les facettes qui exercent l'effet expérimental le plus important : ce sont ces facettes qu'il faudra d'abord contrôler dans un dispositif d'aide au sujet.
2. Le calcul du coefficient de généralisabilité d'erreur absolue permet d'estimer la précision du taux de réussite du sujet ou encore du taux de difficulté des différentes facettes de résolution de problèmes pour des échantillons de sujets tirés de la même population. L'étude d'optimisation permet également de calculer combien de niveaux d'observation de différentes facettes sont nécessaires pour obtenir la précision requise pour situer, par exemple, un sujet à un stade de

développement particulier. Tant le taux de réussite que le taux de difficulté doivent être estimés avec exactitude pour que le diagnostique soit valide.

3. Les développements récents de la théorie de la généralisabilité (Cardinet, 1994) permettent de mesurer la fiabilité du progrès individuel dans différentes conditions au moyen du «*découpage*» de facettes. Cardinet (1994) démontre également qu'il est possible d'estimer non seulement la fiabilité d'un score, mais aussi celle d'un profil de scores. Ces nouveaux ajouts étendent encore davantage les possibilités de l'étude de généralisabilité à la mesure du gain d'apprentissage en situation de mesures répétées ou encore dans l'établissement de profils de scores individuels.

En conclusion, l'utilisation conjointe de la mesure critériée et de la théorie de la généralisabilité permet de mieux mettre à profit l'information obtenue sur la performance d'un sujet ou d'un groupe de sujets à un test, qu'il s'agisse d'un taux de réussite, d'un pourcentage de difficulté ou encore d'un profil de scores. Le diagnostique peut, dans de telles circonstances, être plus fiable et plus valide. La théorie classique des scores comportait des limites sérieuses à cet égard et n'a eu que peu de retentissement dans la mesure opératoire. Les possibilités nouvelles offertes par la théorie de la généralisabilité et la mesure critériée devraient contribuer à renverser cette tendance. Cette article a tenté d'illustrer quelques-unes de ces nouvelles possibilités.

Conclusion

Jacques Grégoire

Après avoir parcouru l'ensemble de l'ouvrage, le lecteur éprouvera peut-être une impression d'hétérogénéité. En fait, cet ouvrage n'a pas été conçu comme un recueil de contributions d'auteurs défendant et illustrant une méthodologie ou un cadre théorique commun. Il représente au contraire un lieu de rencontre et de débat entre des personnes dont l'expérience du terrain et les références théoriques sont différentes. Il vise par conséquent à offrir au lecteur une diversité de réponses à une même question : que peut apporter la psychologie cognitive aux pratiques d'évaluation des apprentissages ? Au travers des chapitres, le lecteur peut ainsi percevoir les forces et les faiblesses, les promesses et les limites des théories cognitivistes dans le champ de l'éducation et, plus particulièrement, dans celui de l'évaluation.

Il nous a paru utile de profiter de ce court chapitre de conclusion pour passer en revue les différentes pratiques d'évaluation des apprentissages et discuter du rôle que peut y jouer la psychologie cognitive. Nous parcourrons ainsi une série de situations qui, chacune, fait appel à un modèle particulier de l'évaluation avec ses exigences et ses règles propres. Ce parcours nous permettra de mieux comprendre la diversité des regards et des jugements que les spécialistes de l'évaluation peuvent porter sur la psychologie cognitive. Dans certaines situations d'évaluation, il est vraisemblable que les apports de la psychologie cognitive resteront limités. Par contre, dans d'autres cadres évaluatifs, la psychologie cognitive stimule déjà des changements importants dans la manière d'apprécier les apprentissages et d'intervenir pour développer ou modifier ceux-ci.

Dans le cadre de l'évaluation certificative, la psychologie cognitive paraît n'avoir guère de place. En effet, la certification porte avant tout sur les performances et non sur les processus mis en oeuvre pour produire ces performances. Aujourd'hui, la préoccupation majeure des praticiens est de construire des épreuves d'évaluation certificative qui soient les plus authentiques possibles, c'est-à-dire qui permettent de vérifier la capacité des apprenants à utiliser adéquatement leurs apprentissages dans des situations de vie quotidienne. Ainsi, l'évaluation certificative de la lecture devrait s'appuyer sur des épreuves fonctionnelles destinées à vérifier si l'élève est capable d'utiliser la lecture en situation naturelle pour atteindre un but précis. Dans une domaine très différent, celui de la certification professionnelle des médecins, des méthodes comme celle du patient standardisé (Swanson, Norman & Linn, 1995) ont été développées pour évaluer la capacité des étudiants à poser un diagnostic correct dans une situation très proche de la situation professionnelle. L'objectif est ainsi de vérifier ce que le candidat médecin est capable de faire et non de mettre en évidence la manière dont il traite l'information qui lui est présentée.

Si nous nous tournons vers l'évaluation en situation de sélection, le constat est assez semblable : la psychologie cognitive n'y a guère de place. Ici aussi, l'attention se porte avant tout sur la performance des sujets, soit en référence à un critère (le sujet atteint-il le seuil de réussite fixé pour être admis ?), soit en référence aux performances des autres candidats (le sujets

se situe-t-il dans les x meilleurs qui seront admis ?). Dans ce cadre évaluatif, les personnes qui présentent les meilleures performances sont réputées posséder certaines caractéristiques cognitives souhaitées. Mais cette inférence reste généralement implicite et ne fait que rarement l'objet d'une analyse approfondie.

Dans le cas de l'évaluation formative, les apports de la psychologie cognitive sont plus diversifiés. Dans un certain nombre de situations, ces apports restent réduits. Ainsi, lors des évaluations en début ou en cours d'apprentissage, de nombreux enseignants se limitent à vérifier la présence ou l'absence de certaines acquisitions. Leur objectif est d'obtenir des informations utiles pour déterminer le point de départ d'une séquence d'apprentissage ou pour décider s'il y a lieu ou non de continuer à exercer telle ou telle habileté. L'attention est alors focalisée sur la réussite des élèves et non sur leur activité mentale. Par contre, certains modèles didactiques encouragent une évaluation formative plus attentive au monde interne des élèves. Ainsi, s'appuyant sur de nombreuses recherches qui montrent que les enfants arrivent à l'école avec des connaissances substantielles à propos des relations quantitatives (voir Bideau, dans ce volume), certains didacticiens suggèrent d'utiliser ces connaissances pour fonder les apprentissages numériques et arithmétiques. Resnick, Bill et Lesgold (1992) encouragent, par exemple, la stimulation des schémas protoquantitatifs que possèdent les enfants (comparaison, croissance/décroissance et partie/tout) pour amener ces derniers à construire des versions quantitatives de ces schémas. Dans cette perspective didactique, des modalités d'évaluation doivent être imaginées pour permettre aux enseignants de réaliser des inférences valides à propos des schémas que possèdent leurs élèves. Un constat similaire peut être fait à propos des apprentissages scientifiques. Pour de nombreux didacticiens des sciences, l'évaluation des préconceptions des élèves apparaît en effet comme une étape nécessaire pour permettre la construction solide de nouvelles connaissances (Juhsua & Dupin, 1993).

La psychologie cognitive peut également éclairer l'évaluation formative via le concept de métacognition (voir Scallon et Wolfs, dans ce volume). On reconnaît aujourd'hui le caractère crucial de l'auto-évaluation pour la construction des apprentissages. Pour apprendre, l'élève doit pouvoir non seulement reconnaître son ignorance mais aussi estimer la qualité de ses connaissances. Par exemple, avant de décider d'utiliser un dictionnaire, un élève doit pouvoir estimer que l'orthographe d'un mot est douteuse. S'il ne peux prendre du recul pour évaluer ses propres productions, il sera incapable de mettre en oeuvre des mécanismes d'autocorrection comme la vérification d'un mot au dictionnaire. Mais, pour aider les élèves à développer des procédures d'autocorrection, il est nécessaire de pouvoir évaluer leurs compétences métacognitives. De ce point de vue, des techniques comme l'évaluation du degré de certitude peuvent être particulièrement utiles (Leclercq, 1975).

La connaissance des contraintes du fonctionnement cognitif a également des prolongements dans l'évaluation formative. Ainsi, la capacité limitée

de la mémoire de travail et la nécessité qui en découle d'automatiser les procédures de bas niveau a des implication non seulement pour l'organisation des apprentissages mais aussi pour la conception de procédures d'évaluation. Fischer (voir supra) présente un bel exemple d'une procédure d'évaluation formative éclairée par la connaissance des mécanismes cognitifs mis en oeuvre dans l'apprentissage des opérations arithmétiques élémentaires. Des procédures d'évaluation formative basées sur des modèles nettement plus complexes du fonctionnement cognitif ont été récemment implantées dans des tutoriels intelligents développés à l'université Carnegie Mellon dans le cadre de l'Advance Computer Tutoring Project (Anderson, 1993). Les résultats obtenus se révèlent très encourageants en terme d'amélioration des apprentissages.

S'il est évident que la psychologie cognitive peut améliorer les pratiques d'évaluation formative, c'est surtout dans le cadre de l'évaluation diagnostique qu'elle se révèle la plus fructueuse. En effet, ce type d'évaluation doit nécessairement déboucher sur une compréhension beaucoup plus précise des processus mis en jeu dans la réalisation d'une tâche. Le simple constat d'une difficulté d'apprentissage nous informe seulement de l'inadéquation entre les caractéristiques cognitives d'un élève et la séquence d'apprentissage qui lui a été proposée. Sur cette base, nous pouvons uniquement conclure que la répétition de cette séquence risque d'être inutile, et même néfaste car il est probable qu'elle entraînera la démotivation de l'élève. Il est donc nécessaire d'aller au-delà du constat d'échec et de tenter de comprendre le sens de la difficulté observée. L'objectif de l'évaluation diagnostique est de mettre en évidence les caractéristiques cognitives des apprenants et, sur cette base, d'élaborer une projet d'action remédiative.

Les procédures d'évaluation diagnostique qui ont été développées dans le cadre de la psychologie cognitive heurtent cependant certains praticiens et chercheurs par leur caractère très analytique et la nature peu authentique des épreuves utilisées. Par exemple, des épreuves de lecture de non mots sont souvent appliquées lors du diagnostic des difficultés en lecture (Jourdain, Zagar & Lété, dans ce volume). Ces épreuve permettent en effet d'évaluer la capacité de segmentation phonologique dont on connaît aujourd'hui le rôle essentiel pour le décodage de mots. La tâche de lecture de non mots est toutefois très peu naturelle. Par ailleurs, elle est issue d'une décomposition de l'activité de lecture en un ensemble d'activités élémentaires. Pour ces deux raisons, elle n'apparaît pas pertinente à certains chercheurs inspirés par les théories de la cognition située *(situated cognition)*.

Pour bien saisir les tenants et les aboutissants du débat dont il est ici question, un bref rappel historique est nécessaire. Comme le souligne Calfee (1995), le behaviorisme n'a pas seulement été mis en question par les travaux cognitivistes sur les caractéristiques internes de l'organisme et sur les processus qui s'y déroulent. D'autres chercheurs se sont intéressés au concept de stimulus et ont souligné les limites de sa définition par les behavioristes. Ainsi, pour Brown, Collins et Duguid (1989) le stimulus doit être défini en terme de contexte. Le contexte est l'ensemble de toutes les circonstances qui affectent

l'individu et qui influencent la production d'un comportement déterminé. Pour ces auteurs, le comportement d'un individu ne peut être compris en dehors du contexte où il est apparu. Selon Calfee (1995), les théories de la cognition située sont complémentaires des théories cognitivistes classiques souvent peu soucieuses du contexte de production des comportements. Toutefois, dans les faits, les tensions sont parfois vives entre ces deux modèles théoriques. Il existe en particulier un courant radical de la cognition située en opposition totale avec les postulats de la psychologie cognitive. Les racines de ce courant radical peuvent être trouvées dans les travaux de Gibson (1979) sur la perception. Selon Gibson, l'information s'offre à nous dans l'environnement et doit simplement être saisie. Il n'est nul besoin de faire appel à un quelconque traitement de cette information. Des notions comme « représentation mentale » et « opération mentale » sont dès lors superflues. Les conceptions de Gibson relatives à la perception ont par la suite été élargies à l'ensemble de la cognition humaine. Dans cette perspective, la compréhension des performances cognitives d'un individu passe par la compréhension de son environnement. Le comportement d'une personne n'est en effet que la réponse directe à la situation rencontrée. Lave (1988) argumente cette conception par les résultats de nombreuses recherches réalisées à propos des mathématiques utilisées par certains enfants et certains adultes dans leur vie quotidienne. Lave rapporte, par exemple, une observation faite sur des clientes de grands magasins. Celles-ci réussissaient 98% des opérations complexes relatives aux produits à acheter (par exemple, *« vaut-il mieux acheter 8 onces de yogourt à 35 cents ou 6 onces de yogourt à 43 cents » ?*). Par contre, ces mêmes personnes ne réussissaient que 70% des mêmes opérations présentées de manière scolaire.

S'appuyant sur de telles observations, les tenants du courant radical de la cognition située défendent la relativité contextuelle de toute activité cognitive. Selon eux, les habiletés d'une personne ne peuvent être comprises qu'en référence au contexte où elles se manifestent. Par ailleurs, ces habiletés ne peuvent être transférées d'un contexte à l'autre et ne peuvent être évaluées hors de leur contexte d'origine. Une telle position est à l'opposé des modèles d'évaluation diagnostique d'inspiration cognitiviste. Pourtant, des cognitivistes comme Anderson (1995) font remarquer qu'une position médiane est tout à fait possible et qu'elle est certainement plus respectueuse des faits. Pour Anderson (1993), une habileté cognitive peut être conçue comme une règle de production de la forme SI-ALORS : si la condition existe alors l'action est exécutée. Plusieurs règles de production peuvent s'organiser autour d'un but et former un système de production. Bien que liées à un contexte donné lors de leur apprentissage, les règles de production ont pour caractéristique de devenir de plus en plus abstraites et de pouvoir être ainsi appliquées à de multiples situations. Par exemple, l'enfant qui a appris à lire son nom en lettres majuscules imprimées peut par la suite le lire dans une infinité de caractères et de contextes de communication. De même, l'élève qui a appris l'algorithme d'addition écrite peut progressivement l'utiliser pour ad-

ditionner une grande diversité de nombres. Petit à petit, cet élève va également lier ce système de production à d'autres systèmes de production et ainsi transférer cette habileté dans de nouveaux contextes. Il ne doit donc pas réapprendre cette habileté pour chaque contexte d'utilisation.

Ce modèle de l'apprentissage est en accord avec un des objectifs fondamentaux de l'enseignement qui est de préparer l'élève à s'adapter à un environnement en perpétuel changement. Pour cela, il est essentiel de favoriser l'abstraction des règles de production et leur transfert dans différents contextes. Dans le cadre de l'évaluation diagnostique des apprentissages, il y a dès lors lieu d'apprécier dans quelle mesure les apprenants transfèrent leurs habiletés cognitives dans une variété de contextes. Une telle évaluation n'est ni contextualisée ni décontextualisée; elle est «transcontextualisée» (Messik, 1994).

Le théorie d'Anderson, que nous venons de présenter très brièvement, permet de surmonter de manière élégante l'opposition entre évaluation décontextualisée et évaluation contextualisée. Elle légitime par ailleurs une analyse des habiletés complexes en leurs composantes élémentaires sans tomber dans le réductionnisme. En ce sens, cette théorie apparaît aujourd'hui comme une cadre de référence prometteur pour l'évaluation diagnostique des apprentissages.

Bibliographie

ALLAL, L. (1979). Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application. In L. ALLAL, J. CARDINET & PH. PERRENOUD (Eds.), *L'évaluation formative dans un enseignement différencié* (pp.10-145). Berne : Peter Lang.

ALLAL, L. (1988). *Peut-on instrumenter l'auto-évaluation ?* Communication présentée lors de la Rencontre de l'ADMEE, Bruxelles.

ALLAL, L. (1988). Vers un élargissement de la pédagogie de maîtrise : Processus de régulation interactive, rétroactive et proactive. In M. HUBERMAN (Ed.), *Assurer la réussite des apprentissages scolaires ? Les propositions de la pédagogie de maîtrise* (pp.86-126). Neuchâtel : Delachaux et Niestlé.

ALLAL, L. (1993). Evaluation formative des processus d'apprentissage : Le rôle des régulations métacognitives. In R. HIVON (Ed.), *L'évaluation des apprentissages.* Sherbroke : Editions du C.R.P.

ALLAL, L. (1993). Régulations métacognitives : quelle place pour l'élève dans l'évaluation formative ? In L. ALLAL, D. BAIN & PH. PERRENOUD (Eds.), *Évaluation formative et didactique du français* (pp.81-98). Neuchâtel : Delachaux et Niestlé.

ANDERSON, J.R. (1982). Acquisition of Cognitive Skill. *Psychological Review, 89*, 369-406.

ANDERSON, J.R. (1983). *The architecture of cognition.* Cambridge, MA : Harvard Univ. Press.

ANDERSON, J.R. (1993). *Rules of the mind.* Hillsdale, NJ : Lauwrence Erlbaum.

ANDERSON, J.R.(1995). *Cognitive psychology and its implications.* New York : Freeman.

ARLIN, P. K. (1975). Cognitive development in adulthood : A fifth stage ? *Developmental Psychology, 11*, 606.

ASHCRAFT, M.H. (1982). The development of mental arithmetic : A chronometric approach. *Developmental Review, 2*, 213-236.

ASHCRAFT, M.H. (1992). Cognitive arithmetic : A review of data and theory. *Cognition, 44*, 75-106.

BADDELEY, A. (1978). The trouble with levels : A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review, 85*, 139-152.

BAIN, D. (1988). L'évaluation formative fait fausse route : De là la nécessité de changer de cap. *Mesure et Evaluation en Education, 10*, 23-32.

BAROODY, A.J. (1991). Procédures et principes de comptage : leur développement avant l'école. In J. BIDEAUD, C. MELJAC & J.P. FISCHER (Eds.), *Les chemins du nombre* (pp.133-158). Lille : PUL

BAROODY, A.J. & GINSBURG, H.P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. HIBERT (Ed.), *Conceptual and procedural knowledge : The case of mathematics* (pp.75-112). Hillsdale, NJ : Lawrence Erlbaum Associates.

BARRON, R. W. (1986). Word recognition in early reading : A review of the direct and indirect access hypotheses. *Cognition*, *24*, 93-119.

BECK, A.T. (1976). *Cognitive therapy and the emotional disorders.* New York : Int. Univ. Press.

BECKER, C. A. (1985). What do we really know about semantic context effects during reading ? In D. BESNER, T.G. WALLER & G.E. MACKINNON (Eds.), *Reading Research : Advances in theory and pratice*, *5*, 125-166.

BEJAR, I.I. & BRAUN, H.I. (1994). On the synergy between assessment and instruction : Early lessons from computer-based simulations. *Machine-Mediated Learning*, *4*, 5-25.

BENTON, S. L. & KIEWRA, K. A. (1985). The Assessment of Cognitive Factors in Academic Abilities. In R.R. RONNING, J. GLOVER, J.C. CONOLEY & J. WITT (Eds). *The Influence of Cognitive Psychology on Testing and Measurement.* Hillsdale, N.J. : Lawrence Erlbaum Associates.

BERKSON, L. (1990). *Critique of Problem-Based Learning : A Review of the Literature.* Unpublished master's thesis, Université of Illinois, Chicago.

BERZONSKY, M. D. (1971). The interdependance of Inhelder and Piaget's model of logical thinking. *Developmental Psychology*, *4*, 469-476.

BIDEAUD, J. (1988). *Logique et bricolage chez l'enfant.* Lille : PUL.

BIDEAUD, J. (1991). Les chemins du nombre. Confrontations et perspectives. In J. BIDEAUD, C. MELJAC & J.-P. FISCHER (Eds.), *Les chemins du nombre* (pp.435-450). Lille : PUL.

BIDEAUD, J. & DAÏF, N. (à paraître). Les significations du nombre dans la vie quotidienne entre 3 et 8 ans.

BIDEAUD, J. & HOUDÉ, O. (1989). Le développement des catégorisations : «capture» logique ou «capture» écologique des propriétés des objets ? *L'année Psychologique*, *89*, 87-123.

BIDEAUD, J. & HOUDÉ, O. (1991). *Cognition et développement.* Berne : Peter Lang.

BIDEAUD, J., HOUDÉ, O. & PÉDINIELLI, J.-L. (1993). *L'Homme en développement.* Paris : P.U.F. (2e éd. 1994).

BIDEAUD, J., MELJAC, C. & FISCHER, J.-P. (1991). *Les chemins du nombre.* Lille : PUL.

BIDEAUD, J. & VILETTE, B. (1995), Les apprentissages numériques élémentaires : Psychogenèse polymorphe et variabilité interindividuelle. In J. LAUTREY (Ed.), *Universel et Individuel en Psychologie.* Paris : P.U.F.

BIGSBY, P. (1988). The visual processor module and normal adult readers. *British Journal of Psychology*, *79*, 455-469.

BONNIOL, J.-J. (1981). Influence de l'explicitation des critères utilisés sur le fonctionnement des mécanismes d'évaluation d'une production scolaire. *Bulletin de Psychologie*, *353*, 173-186.

BRIARS, D. & SIEGLER, R.S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, *20*, 607-618.

BRIEN, R. (1990). *Science cognitive et formation.* Montréal : Presses de l'Université du Québec.

BRISSIAUD, R. (à paraître). Penser l'usage du mot «fois» et l'interaction oral/écrit lors de l'apprentissage initial de la multiplication.

BRISSIAUD, R., CLERC, P. & OUZOULIAS, A. (1992). *J'apprends les maths : CE1. Livre du maître.* Paris : Retz.

BROWN, J.S., & BURTON, R.R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, *2*, 155-192.

BROWN, J.S. & VANLEHN, K. (1980). Repair theory : A generative theory of bugs in procedural skills. *Cognitive Science*, *4*, 379-426.

BROWN, J.S. & VANLEHN, K. (1982). Toward a generative theory of bugs in procedural skills. In T.P. CARPENTER, J. MOSER, & T.A. ROMBERG (Eds.), *Addition and subtraction : A cognitive perspective.* Hillsdale, NJ : Erlbaum.

BROWN, J.S., COLLINS, A. & DUGUID, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32-42.

BRUCK, M. (1990). Word-recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology*, *26*, 439-454.

BRUCKERT, E., MINOW, M. & TETSCHENER, W. (1983). Three-tiered software and VLSI aid developmental system to read text aloud. *Electronics, 56*, 133-138.

BUTLER, R. & NISAN, M. (1986). Effects of No Feedback, Task-Related Comments, and Grades on Intrinsic Motivation and Performance. *Journal of Educational Psychology*, *78*, 210-216.

CALFEE, R. (1995). Implication of cognitive psychology for authentic assessment and instruction. In T. OAKLAND & R.K. HAMBLETON, *International Perpectives on academic assessment* (pp. 25-48). Boston : Kluwer.

CAMPBELL, J.I.D. (1987). Production, verification, and priming of multiplication facts. *Memory & Cognition, 15*, 349-364.

CARDINET, J. & TOURNEUR, Y. (1985). *Assurer la mesure.* Berne : Peter Lang.

CARDINET, J. (1988). *Pour apprécier le travail des élèves.* Bruxelles : De Boeck.

CARDINET, J. (1994). Contrôle de la valeur d'un dispositif intra-individuel. In D. LAVEAULT, B. ZUMBO, M. GESSAROLI & M. BOSS (Eds.), *Modern Theories of Measurement : Problems and Issues* (pp.181-212). Ottawa : Edumetric Research Group.

CARDINET, J. (1986). *Pour apprécier le travail des élèves.* Bruxelles : De Boeck.

CARRAHER, T.N., CARRAHER D.N. & SCHLIEMANN A.D. (1985). Mathematics in the streets and in the school. *British Journal of Developmental Psychology, 3*, 21-29.

CARRAHER, T.N., CARRAHER, D.W. & SCHLIEMANN, A.D. (1987). Written and oral mathematics. *Journal of Research in Mathematics Education, 18*, 89-97.

CASE, R. & SANDIESON , R. (1993). Testing for the presence of a central quantitative structure : use the transfert paradigm. In R. CASE (Ed.), *The mind's staircase.* Hillsdale (N.J.) : Laurence Erlbaum.

CATHCART ,W.G. (1971). The relationship between primary students' rationalization of conservation and their mathematical achievement. *Child Development*, *42*, 755-765.

CHOMSKY, N. (1965). *Aspects of the theory of syntax.* Cambridge. MA : M.I.T. Press.

CLEMENTS, D.H. (1984). Training effects on the development and generalization of piagetien logical operations and knowledge of number. *Journal of Educational Psychology*, *76*, 766-776.

COHEN, R. (Ed.). (1992). *Quand l'ordinateur parle...* Paris : P.U.F.

COLLINS, A. M. & LOFTUS, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

COLTHEART, M. (1978). Lexical access in simple reading tasks. In G. UNDERWOOD (Ed.), *Strategies of information processing* (pp.151-216). New-York : Academic Press.

COLTHEART, M. (1981). Disorders of reading and their implications for models of normal reading. *Visible Language, 15*, 245-286.

CONNOLLY, A.J. (1988). *Key Math Revised : A diagnostic inventory of essentials mathematics*. Circle Pines, MI : American Guidance Service.

CONTENT, A. (1991). The effect of spelling-to-sound regularity on naming in French. *Psychological Research, 53*, 3-12.

CONTENT, A., MOUSTY, P. & RADEAU, M. (1990). Brulex : une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique, 90*, 551-566.

CORDIER, F. (1980). Gradients de prototypie pour cinq catégories sémantiques. *Psychologie Française, 1*, 299-333.

COVINGTON, M. V. & OMELICH, C.L. (1984). Task Oriented Versus Competitive Learning Structure : Motivational and Performance consequences. *Journal of Educational Psychology, 76*, 1038-1050.

CROCKER, L. & ALGINA, J. (1986). *Introduction to Classical and Modern Test Theory.* New York : Holt, Rinehart and Winston.

CRONBACH, L., GLESER, G., NANDA, H. & RAJARATNAM, N. (1972). *The Dependability of Behavioral Measurements : Theory of Generalizability for Scores and Profiles.* New York : Wiley.

CRONBACH, L., RAJARATNAM, N. & GLESER, C. (1963). Theory of generalizability : A liberalization of reliability theory. *British Journal of Mathematical and Statistical Psychology, 16*, 137-163.

CUNNINGHAM, A.E. & STANOVICH, K.E. (1990). Assessing print exposure and orthographic processing skill in children : A quick measure of reading experience. *Journal of Educational Psychology, 82*, 733-740.

DAVIS, R. B. & MCKNIGHT, C.C. (1979). Modeling the processes of mathematical thinking. *Journal of Children's Mathematical Behavior, 2*, 91-113.

DE KETELE, J.M. (1991). Balade au pays de l'auto-évaluation. *Enjeux, 22*, 33-38.

DE LA PASSARDIÈRE, B. & BARON, G.L. (Eds.). (1991). *Hypermédias et apprentissages.* Paris : INRP.

DE VECCHI, G. (1992). *Aider les élèves à apprendre.* Paris : Hachette Education.

DEFRIES, J.C., OLSON, R.K., PENNINGTON, B.F. & SMITH, S.D. (1991). Colorado reading project : An update. In D. DUANE & D. GRAY (Eds.), *The reading brain : The biological basis of dyslexia* (pp.58-87). Parkton, MD : York Press.

DEMERS, C. & TREMBLAY, G. (1992). *Pour une didactique renouvelée de la lecture : du coeur, des stratégies, de l'action...* Rimouski, Québec : Éditions l'Artichaut.

DERRY S. J. & MURPHY D. A. (1986). Designing systems that train learning ability : from theory to practice. *Review of Educational Research, 56*, .

DESMARAIS, M.C., LAROCHELLE, S., & GIROUX, L. (1987). The diagnosis of user strategies. In BULLINGER, H.-J. & SHACKEL (Eds.), Human-Computer Interaction - INTERACT'87. Amsterdam : Elsevier.

DROZ R. (1991). Les multiples racines des nombres naturels et leurs multiples interprétations. In J. BIDEAUD, C. MELJAC & J.-P. FISCHER (Eds.), *Les chemins du nombre* (pp.285-302). Lille : PUL.

DUFFY, S.A. & PISONI, D.B. (1992). Comprehension of synthetic speech produced by rule : A review and theoretical interpretation. *Language and Speech, 35*, 351-389.

DUFOUR, N. (1987). *Guide d'utilisation du logiciel GEN.* Québec : Université Laval, Département de mesure et évaluation.

DULANY, N. (1984). A case of syntactical learning and judgement : How conscious and how abstract ? *Journal of Experimental Psychology, 113*, 541-555.

EHRI, L.C. & ROBBINS, C. (1992). Beginners need some decoding skill to read words by analogy. *Reading Research Quarterly, 27*, 13-26.

EHRLICH, S. & FLORIN, A. (1989). Ne pas décourager l'élève : Etude sur l'échec de fonctionnement des enfants en classe. *Revue Française de Pédagogie, 86*, 35-48.

ELKIND, J., COHEN, K. & MURRAY, C. (1993). Using computer-based readers to improve reading comprehension of students with dyslexia. *Annals of Dyslexia, 43,* 238-259.

ELLIS, A. (1962). *Reason and emotion in psychotherapy.* Secaucus, NJ : Citadel Press.

ELLIS, A.W. (1989). *Lecture, écriture et dyslexie. Une approche cognitive.* Neuchâtel : Delachaux & Niestlé.

ELLIS, A. W., & YOUNG, A. W. (1990). *Human Cognitive Neuropsychology.* Hillsdale, NJ : Lawrence Erbaum.

ERICSSON, K. A. (1985). Theoretical Implications from Protocol Analysis on Testing and Measurement. In R. R. RONNING, J. GLOVER, J. C. CONOLEY & J. C. WITT (Eds), *The Influence of Cognitive Psychology on Testing and Measurement.* Hillsdale, N.J. : Lawrence Erlbaum Associates.

EVETT, L. J. & HUMPHREYS, G. W. (1981). The use of abstract graphemic information in lexical access. *Quaterly Journal of Experimental Psychology, 33*, 325-350.

FARMER, M.E., KLEIN, R. & BRYSON, S.E. (1992). Computer-assisted reading : Effects of whole-word feedback on fluency and comprehension in readers with severe disabilities. *Remedial and Special Education, 13,* 50-60.

FAYNE, H.R. & BRYANT, N.D. (1981). Relative effects of various word synthesis strategies on the phonics achievement of learning disabled youngsters. *Journal of Educational Psychology, 73*, 616-623.

FAYOL, M. (1990). *L'enfant et le nombre.* Neuchâtel : Delachaux & Niestlé.

FISCHER, J.-P. (1987). L'automatisation des calculs élémentaires à l'école. *Revue Française de Pédagogie, 80*, 17-24.

FISCHER, J.-P. (1988a). *11-3=9 : Juste ou Faux ? Une méthode moderne d'évaluation de - et des progrès dans - la connaissance des faits numériques élémentaires.* Montigny-lès-Metz : CDDP de la Moselle.

FISCHER, J.-P. (1988b). La mesure des TR en arithmétique élémentaire : Spécificités d'une tâche de vérification. In R. DUVAL (Ed.), *Annales de Didactique et de Sciences Cognitives* (pp.153-178). Strasbourg : IREM.

FISCHER, J.-P. (1992). *Apprentissages numériques : la distinction procédural/ déclaratif.* Nancy : Presses Universitaires.

FISCHER, J.-P. (1993a). De quelques notions-clés concernant l'acquisition du nombre. *Faits de Langues, 2*, 7-16.

FISCHER, J.-P. (1993b). *Trois grands théoriciens des apprentissages scolaires.* Strasbourg : IREM.

FISCHER, J.-P. & PLUVINAGE, F. (1988). Complexités de compréhension et d'exécution des opérations arithmétiques élémentaires. *Recherches en Didactique des Mathématiques, 9*, 133-154.

FLAVELL, J.H. (1979). Metacognition and cognitive monitoring : A new area of cognitive developmental inquiry. *American Psychologist, 34*, 906-911.

FONTEYN, A. (1993). *Approche métacognitive des méthodes de travail citées par des étudiants de première candidature.* Mémoire de licence inédit, Université libre de Bruxelles, Bruxelles.

FORSTER, K. I. (1976). Accessing the mental lexicon. In R.J. WALES & E. WALKER (Eds.), *New approaches to language mechanisms* (pp.257-288). Amsterdam : North-Holland.

FOX, B. & ROUTH, D. (1980). Phonemic analysis and severe reading disability in children. *Journal of Psycholinguistic Research, 9*, 115-119.

FRENCK-MESTRE, C. & VAID, J. (1993). Activation of number facts in bilinguals. *Memory & Cognition, 21*, 809-818.

FUSON , C.K., RICHARD, J. & BRIARS, D.J. (1982). The acquisition and elaboration of the number word sequence. In C. BRAINERD (Ed.), *Progress in cognitive development : Children's logical and mathematical cognition,* (Vol.1, pp.33-92) New York : Springer-Verlag.

FUSON, C.K. (1988). *Children's counting and concepts of number.* New York : Springer-Verlag.

FUSON, C.K. (1991). Relations entre comptage et cardinalité chez les enfants de 2 à 8 ans. In J. BIDEAUD, C. MELJAC ET J.-P. FISCHER (Eds.), *Les chemins du nombre* (pp.159-179). Lille : Presses Universitaires de Lille.

FUSON, C.K., SECADA ,W.G. & HALL, J.W. (1983). Matching counting and conservation of numerical equivalence. *Child Development, 54*, 91-97.

GAGNE, R.M. (1984). Learning Outcomes and Their Effects : Useful Categories of Human Performance. *American Psychologist, 39*, 377-385.

GARDNER, H. (1985). *The mind's new science. A history of the cognitive revolution.* New York : Basic Books.

GEARY, D.C. (1993). Mathematical disabilities : Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114*, 345-362.

GEARY, D.C., BROWN, S.C. & SAMARANAYAKE, V.A. (1991). Cognitive addition : A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology, 27*, 787-797.

GEARY, D.C., FRENSCH P.A. & WILEY, J.G. (1993). Simple and complex mental subtraction : Strategy choice and speed-of-processing differences in younger and older adults. *Psychology and Aging, 8*, 242-256.

GEARY, D.C. & WILEY, J.G. (1991). Cognitive addition : Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging, 6*, 474-483.

GELMAN, R. & GALLISTEL, C.R. (1978). *The child understanding of number.* Cambridge, MA : Harvard University Press.

GELMAN, R. & MECK, E. (1983). Preschooler's counting : Principles before skill. *Cognition, 13*, 343-359.

GELMAN, R. & MECK, E. (1991). Premiers principes et conceptions du nombre. In J. BIDEAUD, C. MELJAC ET J.-P. FISCHER (Eds.), *Les chemins du nombre* (pp.211-234) Lille : Presses Universitaires de Lille.

GELMAN, R. & TUCKER, M.F. (1975). Futher investigation of the young child's conception of number. *Child Development, 46*, 167-173.

GELMAN, R. (1972). The nature and development of early number concepts. In H.W. REESE (Ed.), *Advances in child development and behavior* (pp.115-167). New York : Academic Press.

GIBSON, J.J. (1979). *The ecological approach to visual perception.* Boston : Houghton-Mifflin.

GINSBURG, H.P. & RUSSEL, R.L. (1981). Social class and racial influences on early mathematical thinking. *Monographs of the Society for Research in Child Development, 46*, 1-67.

GLASER, R. (1986). Enseigner comment penser : Le rôle de la connaissance. In M. CRAHAY & D. LAFONTAINE (Eds.), *L'art et la science de l'enseignement.* Bruxelles : Labor.

GLASER, R., LESGOLD, L. & LAJOIE, S. (1985). Toward a Cognitive Theory for the Measurement of Achievement. In R. R. RONNING, J. GLOVER, J. C. CONOLEY ET J. C. WITT (Eds), *The Influence of Cognitive Psychology on Testing and Measurement* (pp.41-85). Hillsdale, N. J. : Lawrence Erlbaum Associates.

GLASS, G.V. (1985). Testing Old, Testing New : Schoolboy Psychology and the Allocation of Intellectual Resources. In B. PLAKE & J. C. WITT (Eds), *The Future of Testing* (pp.9-27). Hillsdale, N.J. : Lawrence Erlbaum Associates.

GOMBERT, J.E. (1990). *Le développement métalinguistique.* Paris : PUF.

GRAY, W. M. (1978). A comparison of piagetian theory and criterion-referenced measurement. *Review of Educational Research, 48*, 223-249.

GRÉCO, P. & MORF, A. (1962). *Structures numériques élémentaires (Etudes d'Epistémologie génétique, Vol. XIII).* Paris : PUF.

GRÉCO, P. (1962). Quantité et quotité. In P. GRÉCO ET A. MORF (Eds.), *Structures numériques élémentaires.* Paris : PUF.

GRÉCO, P. (1962). Une recherche sur la commutativité de l'addition. In P. GRÉCO & A. MORF (Eds.), *Structures numériques élémentaires* (pp.1-70). Paris : PUF.

GREENE, B.G., LOGAN, J.S. & PISONI, D.B. (1986). Perception of synthetic speech produced by automatically by rule : Intelligibility of eight text-to-speech systems. *Behavior Research Methods, Instruments, & Computers, 18,* 100-107.

GREENSPAN, S.L., NUSBAUM, H.C. & PISONI, D.B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 14*, 421-433.

GRÉGOIRE, J. & VAN NIEUWENHOVEN, C. (1994). Le comptage en troisième maternelle et en première primaire : un outil d'évaluation précoce, *Cahiers du Laboratoire de Pédagogie expérimentale de l'Université de Louvain, 7*, 67-90.

GRÉGOIRE, J. (1992). Evaluer les troubles cognitifs au moyen des épreuves piagétiennes ? Analyse de quelques problèmes méthodologiques. *Archives de Psychologie, 60*, 177-204.

GROEN, G. & PARKMAN, J. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329-343.

GROEN, G. & RESNICK, L.B. (1977). Can preschool children invent addition algorithms ? *Journal of Educational Psychology, 69*, 645-662.

GUTTMAN, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*, 139-150.

HASEMANN, K. (1985). Analysis of fraction errors by a model of cognitive science. *European Journal of Psychology of Education, 1*, 57-66.

HATON, M.C. (1992). L'ordinateur pédagogue. *La Recherche, 23*, 1014-1022.

HOUDÉ, O. (1993). Pensée logico-mathématique et psychologie. In O. HOUDÉ & D. MIÉVILLE (Eds.), *Pensée logico-mathématique* (pp.47-119). Paris : PUF.

HOWELL, R., SIDORENKO, E. & JURICA, J. (1987). The effects of computer use on the acquisition of multiplication facts by a student with learning disabilities. *Journal of Learning Disabilities, 20*, 336-341.

INHELDER, B., SINCLAIR, H. & BOVET, M. (1974). *Apprentissage et structure de la connaissance.* Paris : PUF.

INHELDER, B. & PIAGET, J. (1955). *De la logique de l'enfant à la logique de l'adolescent.* Paris : PUF.

INHELDER, B. (1943). *Le diagnostic du raisonnement chez les débiles mentaux.* Neuchâtel : Delachaux et Niestlé, (2ème éd., 1969).

JANVIER, C. (1991). Contextualisation et représentation dans l'utilisation des mathématiques. In C. GARNIER, N. BEDNARZ & I. ULANOVSKAYA (Eds.), *Après Vygotski et Piaget* (pp.129-147). Bruxelles : De Boeck.

JASTAK, S. & WILKINSON, G.S. (1984). *Wide Range Achievement Test-Revised.* Wilmington, DE : Jastak Associates.

JONNAERT, PH. (1994). *L'enfant-géomètre.* Bruxelles : Plantyn.

JORM, A.F. & SHARE, D.L. (1983). Phonological recording and reading acquisition. *Applied Psycholinguistics, 4*, 103-147.

JOHSUA, S. & DUPIN, J.-J. (1993). Introduction à la didactique des sciences et des mathématiques. Paris : PUF.

JOURDAIN, C. (1989). Un projet d'évaluation-diagnostic des difficultés en lecture de jeunes adultes. Rapport de DESS non publié, Universite de Bourgogne, Dijon.

KAHNEMAN, D. (1973). *Attention and effort.* Englewood Cliffs, NJ : Prentice-Hall.

KAMII, C.K. (1985). *Young children reinvent arithmetic : implications of Piaget's theory*, New-York : Teacher's college, Colombia University.

KAMII, C. (1990). *Les jeunes enfants réinventent l'arithmétique.* Berne : Peter Lang

KARPLUS, R. & PETERSON, R. W. (1970). Intellectual development beyond elementary school II : Ratio, a survey. *School Science and Mathematics, 9*, 813-820.

KARPLUS, R. (1977). Science teaching and the development of reasoning. *Journal of Reasoning in Science Teaching, 14*, 169-175.

KAYE, D.B. (1986). The development of mathematical cognition. *Cognitive Development, 1*, 157-170.

KITCHENER, K.S. (1983). A three Level model of Cognitive Processing. *Human development, 26*, 222-232.

KLATT, D.H. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America, 82*, 738-793.

KLUGER, A.N. & ADLER, S. (1993). Person- versus computer-mediated feedback. *Computers in Human Behavior, 9*, 1-16.

KRAIGER, K.J., FORD, K. & SALAS, A. (1993). Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. *Journal of Applied Psychology (Monograph), 78*, 311-328.

LACHMAN, R., LACHMAN, J.L. & BUTTERFIELD, E.C. (1979). *Cognitive Psychology and information processing : An introduction.* Hillsdale, NJ : Lawrence Erlbaum.

LABERGE, D. & SAMUELS, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293-323.

LANKFORD, F. G. (1972). *Some computational strategies of seventh grade pupils* (Final report, Project No. 2-C-013). HEW/OE National Center for Educational Research and Development and the Center for Advanced Studies, University of Virginia.

LARIVÉE, S., NORMANDEAU, S., ROULIN, J.-L. & LONGEOT, F. (1987) L'épreuve de la balance de Siegler : Analyse critique du modèle par élaboration de règles. *L'Année Psychologique*, *87*, 509-534.

LAUTREY, J. (1980). *Classe sociale, milieu familial et intelligence.* Paris : PUF.

LAUTREY, J., DE RIBEAUPIERRE, A. & RIEBEN, L. (1981). Le développement opératoire peut-il prendre des formes différentes chez des enfants différents ? *Journal de Psychologie Normale et Pathologique*, *4.*, *421-443.*

LAVE, J. (1988). *Cognition in practive : Mind, mathematics, and culture in everyday life.* New York : Cambridge University Press.

LAVEAULT, D. (1987). *Test Imageo5D d'habiletés spatiales.* Ottawa : Faculté d'éducation, Université d'Ottawa.

LAVEAULT, D. (1990). Evaluation par objectifs, une approche métacognitive. *Mesure et Evaluation en Education*, *13*, 57-74.

LAWSON, A. E. (1974). Piagetian formal operational tasks : A crossover study of learning effect and reliability. *Science Education*, *58*, 267-274.

LECLERCQ, D. (1991). Peut-on estimer l'auto-estimation des compétences dans la pratique scolaire ? Qu'en aurait dit Descartes ? *Enjeux*, *22*, 64-73.

LECLERCQ, D. (1975). *L'évaluation subjective de la probabilité d'exactitude des réponses en situation pédagogique.* Thèse de doctorat non publiée. Université de Liège, Liège.

LEISER, D. & GILLIERON, CH. (1990). *Cognitive science and genetic psychology.* New York : Plenum Press.

LEMAIRE, P. (1993). *La récupération des faits arithmétiques élémentaires chez des adultes et des enfants. De l'acquisition à la modélisation.* Thèse de Doctorat non publiée, Université de Bourgogne, Dijon.

LEMAIRE, P., BARRETT, S.E., FAYOL, M. & ABDI, H. (1994). Automatic activation of addition and multiplication facts in elementary school children. *Journal of Experimental Child Psychology*, *57*, 224-258.

LEMAIRE, P., FAYOL, M. & ABDI, H. (1991). Associative confusion effect in cognitive arithmetic : Evidence for partially autonomous processes. *Cahiers de Psychologie Cognitive / European Bulletin of Cognitive Psychology*, *11*, 587-604.

LEONG, C.K. (1992A). Enhancing reading comprehension with text-to-speech (DECtalk) computer system. *Reading and Writing : An interdisciplinary Journal*, *4*, 205-217.

LEONG, C.K. (1992B). Text-to-speech, text, and hypertext : Reading and spelling with the computer. *Reading and Writing : An interdisciplinary Journal*, *4*, 95-105.

LEONG, C.K. (1993). Towards an applied cognitive science perspective in education. *International Journal of Disability, Development and Education*, *40*, 63-73.

LESELBAUM, N. (1982). *Autonomie et auto-évaluation.* Paris : Economica.

LÉTÉ, B., ZAGAR, D. & PYNTE, J. (1994). A new paradigm for studying word recognition during reading. Exploration with word frequency effects. *International Journal of Psychology, 29*, 291-318.

LINN, M. C. & LEVINE, D. I. (1977). Scientific reasoning in adolescence : Theoretical viewpoints and educational implications. *Journal of Research in Science Teaching, 12*, 49-62.

LOGAN, J.S., GREENE, B.G. & PISONI, D.B. (1989). Segmental intelligibility of synthetic speech produced by rule. *Journal of the Acoustical Society of America, 86*, 566-581.

LONA, A. (1994). *Comment les étudiants prennent-ils note ?* Mémoire de licence non publié, Université de Mons-Hainaut, Mons.

LONDEIX, H. (1985). Les relations entre psychologie génétique et psychologie différentielle. I. Aperçus sur les objectifs et les méthodes. *L'orientation scolaire et professionnelle, 14*, 23-43.

MANDLER J. M. (1979). Categorical and schematic organization in memory. In C.R. PUFF (Ed.), *Memory organization and structure* (pp.259-299). New York : Academic Press.

MANOUS, L.M. & PISONI, D.B. (1984). Effects of signal duration on the perception of natural and synthetic speech. *Research on Speech Perception Progress Report, 10*, 311-321.

MARTORANO, S. C. (1977). A developmental analysis of performance on Piaget's formal operations tasks. *Developmental Psychology, 13*, 666-672.

MAYER, R.E. (1985). Mathematical ability. In R.J. STERNBERG (Ed.). *Human abilities : An information-processing approach* (pp.127-140). New York : Freeman & Co.

MCCLOSKEY, M. (1992). Cognitive mecanismes in numerical processing : Evidence from acquired dyscalculia. *Cognition, 44*, 107-157.

MCCLELLAND, J.L. & RUMELHART, D.E. (1981). An interactive activation model of context effects in letter perception during reading : Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

MCCOLSKEY, W. & LEARY, M.R. (1985). Differential Effects of Norm-Referenced and Self-Referenced Feedback on Performance, Expectancies, Attributions, and Motivation. *Contemporary Educational Psychology, 10*, 275-284.

MCCONKIE, G.W. & ZOLA, D. (1985). *Computer aided reading : An environment for developmental research.* Paper presented at the meeting of the Society for Research in Child Development, Toronto.

MCCUSKER, L.X., HILLINGER, M. L. & BIAS, R.G. (1981). Phonological recoding and reading. *Psychological Bulletin, 89*, 217-245.

MELJAC, C. (1980). *Batterie UDN 80 : Construction et utilisation des premiers nombres*. Paris : ECPA.

MELJAC, C.(1988). L'examen opératoire. In P. MAZET & S. LEBOVICI (Eds), *Penser, apprendre* (pp.119-209). Paris : Eshel.

MELJAC, C. (1991). De quelques variantes imprévues apportées au scénario de la construction du nombre, in J. BIDEAUD, C. MELJAC ET J.-P. FISCHER (Eds.), *Les chemins du nombre* (pp.401-432). Lille : PUL.

MEN, J. (1991). *Les cycles à l'école primaire.* Paris : Hachette.

MESSICK, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*, 13-23.

MEYER, D. E., SCHVANEVELDT, R. W. & RUDDY, M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. *Memory & Cognition, 2*, 309-321.

MICHIE, S. (1984). Why preschoolers are reluctant to count spontaneously. *The British Journal of Developmental Psychology, 2*, 347-358.

MILLER, K., PERLMUTTER, M. & KEATING, D. (1984). Cognitive arithmetic : Comparison of operations. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 10*, 46-60.

MINISTÈRE DE L'EDUCATION (1979). *Programme d'études primaires. Français.* Québec : Gouvernement du Québec.

MINSKY, M. (1975). A framework for representing knowledge. In P. WINSTON (Ed.), *The psychology of computer vision* (pp.211-277). New York : McGraw-Hill.

MITCHELL, D. C. (1987). Lexical guidance in human parsing : locus and processing characteristics. In M. COLTHEART (Ed.), *Attention and Performance XII : The Psychology of Reading* (pp.601-618). London, Hillsdale : Lawrence Erbaum Associates.

MOUNOUD, P. (1994). L'émergence de conduites nouvelles : rapports dialectiques entre systèmes de connaissances. *Psychologie & Education, 18*, 11-42.

NEIMARK, E. D. (1975). Longitudinal development of formal operational thought. *Genetic Psychology Monographs, 91*, 171-226.

NEISSER, U. (1994). Multiple systems : A new approach to cognitive theory. *European Journal of Cognitive Psychology, 6*, 225-241.

NOËL, B., ROMAINVILLE, M. & WOLFS, J.-L. (à paraître). La métacognition : facettes et pertinence du concept en éducation. *Revue Française de Pédagogie.*

NOËL, B.(1988). *Contribution à l'étude de la métacognition (Elaboration d'un modèle à partir d'analyse de cas d'enfants de fin d'école primaire).* Thèse de doctorat non publiée, Université de Mons-Hainaut, Mons.

NOËL, B. (1991). *La métacognition.* Bruxelles : De Boeck.

NUNES, T., SCHLIEMANN, A.D. & CARRAHER, D.W. (1993). *Street mathematics and school mathematics.* Cambridge : Cambridge Univ. Press.

NUNZIATI, G. (1990). Pour construire un dispositif d'évaluation formative. *Cahiers Pédagogiques, 280*, 47-64.

NYITI, R. M. (1982). The validity of «cultural differences explanations» for cross-cultural variation in the rate of Piagetian cognitive development. In D. A. WAGNER & H. W. STEVENSON (Eds.), *Cultural Perspectives on Child Development.* San Francisco : W. H. Freeman and Co.

OLOFSSON, A. (1992). Synthetic speech and computer aided reading for disabled children. *Reading and Writing : An interdisciplinary Journal, 4,* 165-178.

OLSON, R. K., KLIEGL, R., DAVIDSON, B. J. & FOLTZ, G. (1985). Individual and developmental differences in reading disability. In G. E. MACKINNON & T. G. WALLER (Eds.), *Reading research : Advances in theory and pratice.* New-York : Academic Press.

OLSON, R.K., FOLTZ, G. & WISE, B.W. (1986). Reading instruction and remediation with the aid of computers speech. *Behavior Research Methods, Instruments, & Computers, 18,* 93-99.

OLSON, R.K. & WISE, B.W. (1992). Reading on the computer with orthographic and speech feedback. *Reading and Writing : An interdisciplinary Journal, 4,* 107-144.

OLSON, R.K., WISE, B. & RACK, J.P. (1989). Dyslexia : Deficits, genetic aetiology, and computer-based remediation. *The Irish Journal of Psychology, 10,* 494-508.

OREN, D. L. (1983). Evaluation Systems and Attributional Tendencies in the Classroom : A Sociological Approcah. *Journal of Educational Research, 76*, 307-312.

PAAP, K. R., NEWSOME, S. L., MCDONALD, J. E. & SCHVANEVELDT, R. W. (1982). An activation-verification model for letter and word recognition : the word-superiority effect. *Psychological Review, 89*, 573-594.

PAPERT, S. (1993). *The children's machine. Rethinking school in the age of computer.* New York : Basic Books.

PAQUAY L., ALLAL, L. & LAVEAULT, D. (1990). L'auto-évaluation en question(s) : Propos pour un débat. *Mesure et Evaluation en Education, 13*, 5-26.

PENNINGTON, B.F., WALLACH, L. & WALLACH, M.A. (1980). Nonconservers' use and understanding of number and arithmetic. *Genetic Psychology Monographs, 101*, 231-243.

PERKINS, D. N. & SALOMON, G. (1989). Are Cognitive Skills Context-Bound ? *Educational Researcher, 18*, 16-25.

PIAGET, J. (1947). *La psychologie de l'intelligence.* Paris : Armand Collin (8ème éd., 1967).

PIAGET, J.(1974). *La prise de conscience.* Paris : PUF.

PIAGET, J., HENRIQUES, G. & ASCHER, E. (1990). *Morphismes et catégories : Comparer et transformer.* Neuchâtel : Delachaux et Niestlé.

PIAGET, J. & SZEMINSKA, A. (1941). *La genèse du nombre chez l'enfant.* Neuchâtel : Delachaux et Niestlé, (7ème éd., 1991).

PISONI, D.B. (1981). Speeded classification of natural and synthetic speech in a lexical decision task. *Journal of the Acoustical Society of America, 70*, 898.

POPHAM, W. J. & HUSEK, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement, 6*, 1-9.

POSNER, M. I. & MITCHELL, D. C. (1967). Chronometric analysis of classification. *Psychological Review, 74*, 392-409.

RANSCHBURG, P. (1930). Experimentelle Beiträge zur Lehre von Gedächtnis, Urteil und Schlussfolgerung an Gesunden und Kranken. *Archiv für die gesamte Psychologie, 77*, 437-526.

RAYNER, K. & POLLATSEK, A. (1987). Eye movements in reading : a tutorial review. In M. COLTHEART (Ed.), *Attention and Performance XII : The Psychology of Reading* (pp.327-362). London, Hillsdale : Lawrence Erbaum Associates.

RAYNER, K. & POLLATSEK, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ : Prentice Hall.

RESNICK, L.B. & RESNICK, D.P. (1990). Assessing the thinking curriculum : New tools for eductional reform. In R.B. GIFFORD & M.C. O'CONNOR (Eds.), *Changing assessments : Alternative views of aptitude, achievement and instruction* (pp.37-75). Boston : Kluwer.

RESNICK, L.B., BILL, V. & LESGOLD, S. (1992). Developing thinking abilities in arithmetic class. In A. DEMETRIOU, M. SHAYER & A. EFKLIDES (Eds.), *Neo-piagetian theories of cognitive development* (pp. 210-230). London : Routledge.

RETSCHITZKI, J. (1990). *Stratégies des joueurs d'awélé.* Paris : L'Harmattan.

REUCHELIN, M. (1991). *Les différences individuelles à l'école.* Paris : PUF.

RIBAUPIERRE, A. (1983). Un modèle piagétien du développement : la théorie des opérateurs constructifs de Pascual-Leone. *Cahiers de Psychologie Cognitive, 3*, 327-356.

RICHELLE, M. (1993). *Quoi de neuf sur l'esprit ?* Paris : PUF.

RIEUNAUD, J. (1989). *L'approche du nombre par le jeune enfant.* Paris : PUF.

RILEY, M.S., GREENO, J.G. & HELLER, J.I. (1983). Development of children's problem-solving ability in arithmetic. In H. GINSBURG (Ed.), *The developement of mathematical thinking* (pp.153-196). New York : Academic Press.

ROBERGE, J. J. (1976). Developmental analyses of two formal operational structures : Combinatorial thinking and conditional reasoning. *Developmental Psychology, 12*, 563-564.

ROMAINVILLE, M. (1992). *Savoir parler de ses manières d'apprendre.* Thèse de doctorat non publiée, Université Catholique de Louvain, Louvain-la-Neuve.

ROID, G.H. (1982). *A technology for test-item writing.* New York : Academic Press.

ROMAINVILLE, M. (1993). *Savoir parler de ses méthodes.* Bruxelles : De Boeck.

ROYER, J.R., CISERO, C.A. & CARLO, M.S. (1993). Techniques and Procedures for Assessing Cognitive Skills. *Review of Educational Research, 63*, 201-243.

ROZIN, P. & GLEITMAN, L.R. (1977). The structure and acquisition of reading, II : The reading process and the acquisition of the alphabetic principle. In A.S. REBER & D. SCARBOROUGH (Eds.), *Toward a psychology of reading* (pp. 55-141). Hillsdale, NJ : Erlbaum.

SALOMON, G. (1993). *Distributed cognitions. Psychological and educational considerations.* Cambridge : Cambridge Univ. Press.

SALTHOUSE, T. (1984). L'art de la dactylographie. *Pour la Science, 78*, 32-37.

SCALLON, G. (1988a). *L'évaluation formative des apprentissages. La réflexion.* Québec : Presses Universitaires de Laval.

SCALLON, G. (1988b). *L'évaluation formative des apprentissages. L'instrumentation.* Québec : Presses Universitaires de Laval.

SCALLON, G. (1981). *La construction d'un test diagnostique selon des facettes, partie I : présentation d'un modèle de recherche pédagogique.* Ste-Foy : Département de mesure et évaluation, Université Laval.

SCALLON, G. (1982). *La construction d'un test diagnostique selon des facettes, partie II : exemples d'application et aspects méthodologiques.* Ste-Foy : Département de mesure et évaluaiton, Université Laval.

SCALLON, G. (1992). *L'impact des pratiques évaluatives sur la motivation des élèves : le pouls de la recherche.* Département de mesure et évaluation, Université de Laval.

SCALLON, G. (1992). «L'évaluation formative : entre la docimologie et la didactique.». In D. LAVEAULT (Ed.), *Les pratiques d'évaluation en éducation.* Montréal, Québec : Éditions de l'ADMÉÉ.

SCALLON, G. (1993). «Tendances actuelles en évaluation formative.» Exposé tenu dans le cadre du symposium : *L'évaluation formative sans examen : un changement à amorcer.* 16e Session d'étude de l'ADMÉÉ, Montréal, 29 octobre 1993.

SCARDAMALIA, M., BEREITER, C., MCLEAN, R., SWALLOW, J. & WOODRUFF, E. (1989). Computer-supported intentional learning environments. *Journal of Educational Computing Research, 5*, 51-68.

SCHVANEVELDT, R. W., ACKERMAN, B. P. & SEMLEAR, T. (1977). The effect of semantic context on childrens word-recognition. *Child Development*, *48*, 612-616.

SCHWAB, E.C., NUSBAUM, H.C. & PISONI, D.B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*, 395-408.

SEGUI, J. (1991). La reconnaissance visuelle des mots. In R. KOLINSKY, J. MORAIS & J. SEGUI (Eds.), *La reconnaissance des mots dans différentes modalités sensorielles. Données et modèles en psycholinguistique cognitive* (pp.99-117). Paris : P.U.F.

SEGUI, J., MEHLER, J., FRAUENFELDER, U. & MORTON, J. (1982). The word frequency effect and lexical access. *Neuropsychologica*, *6*, 615-627.

SERON, X. (1991). Du diagnostic neuropsychologique à l'évaluation cognitive et pragmatique des troubles. *Revue Suisse de Psychologie*, *50*, 186-197.

SERON, X. & DELOCHE, G.(1994). Les troubles du calcul et du traitement des nombres. In X. SERON & M. JEANNEROD (Eds.), *Neuropsychologie humaine* (pp.439-452). Liège : Mardaga

SEYMOUR, P. H. K. (1986). *Cognitive analysis of dyslexia.* London : Routledge & Kegan.

SIEGLER, R.S. (1988). Individual differences in strategy choices : good students, not-so good students and perfectionists. *Child Development*, *59*, 833-851.

SIEGLER, R.S. (1987). The perils of averaging data over strategies : An example from children's addition. *Journal of Experimental Psychology : General*, *116*, 250-264.

SIEGLER, R.S. (1989). Hazards of mental chronometry : An example from children's subtraction. *Journal of Educational Psychology*, *81*, 497-506.

SIEGLER, R.S., & SHRAGER, J. (1984). Strategy choices in addition and subtraction : How do children know what to do ? In C. SOPHIAN (Ed.), *Origins of cognitive skills* (pp. 229-293). Hillsdale : Erlbaum.

SMITH, F. (1971). *Understanding reading : A psycholinguistic analysis of reading and learning to read.* New-york : Holt, Rinehart & Winston.

SMITHER, J.A. (1993). Short-term memory demands in processing synthetic speech by old and young adults. *Behaviour and Information Technology, 12,* 330-335.

SOPHIAN, C. (1987). Early developments in children's use of counting to solve quantitative problems. *Cognition and Instruction*, *4*, 61-90.

SOPHIAN, C. (1988). Limitations on children's knowledge about counting : using counting to compare two sets. *Developmental Psychology*, *24*, 634-640.

SOPHIAN, C. (1991). Le nombre et sa genèse avant l'école primaire. Comment s'en inspirer pour enseigner les mathématiques. In J. BIDEAUD, C. MELJAC ET J.-P. FISCHER (Eds.), *Les chemins du nombre* (pp.35-58). Lille : PUL.

SPAAI, G.W.G., REITSMA, P. & ELLERMANN, H.H. (1991). Effects of segmented and whole-word sound feedback on learning to read single words. *Journal of Educational Research, 84*, 204-213.

SPOEHR, K.T. & SMITH, E.E. (1973). The role of syllables in perceptual processes. *Cognitive Psychology, 5*, 71-89.

STANOVICH, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32-71.

STANOVICH, K.E. (1982). Individual differences in the cognitive processes of reading : 1. Word decoding. *Journal of Learning Disabilities*, *15*, 485-493.

STANOVICH, K.E. (1986). Explaining the variances in reading ability in terms of psychological processes : What have we learned ? *Annals of Dyslexia, 35*, 67-96.

STANOVICH, K.E. & WEST, R.F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24,* 402-433.

STARKEY, P. & GELMAN, R. (1982). The development of addition and substraction abilities prior to formal schooling in arithmetic. In T.P. CARPENTER, J.M. MOSER & T.A. ROMBERG (Eds.), *Addition and substraction : A cognitive perspective* (pp.99-116). Hillsdale, NJ : Lawrence Erlbaum Associates.

STEFFE, L.P. & COBB, P. (1988). *Construction of arithmetical meanings and strategies.* New York : Springer Verlag.

STEFFE, L.P. (1991). Stades d'apprentissage dans la construction de la suite des nombres. In J. BIDEAUD, C. MELJAC ET J.-P. FISCHER (Eds.), *Les chemins du nombre* (pp.113-132). Lille : PUL.

STERN, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children ? *Journal of Educational Psychology, 85*, 7-23.

STERNBERG, R.J. & WAGNER, R.K.(1984). Alternative conceptions of intelligence and their implications for Education. *Review of Education Research, 54*, 179-223.

STERNBERG, R.J. (1990). T & T is an explosive combination : Technology and testing. *Educational Psychologist, 25*, 201-222.

STERNBERG, S. (1969). The discovery of processing stages : extensions of Donder's method. In W.G. KOSTER (Ed.), Attention and Performance II. *Acta Psychologica, 30*, 276-315.

STROOP, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 6*, 501-531.

SWANSON, D.B., NORMAN, G.R. & LINN, R.L. (1995). Performance-based assessment : lessons from health professions. *Educational Researcher, 24*, 5-11.

TAFT, M. (1979). Lexical access via an orthographic code : The Basic Orthographic Syllabic Structure (BOSS). *Journal of Verbal Learning and Verbal Behavior, 18,* 21-39.

TARDIF, J. (1992). *Pour un enseignement stratégique. L'apport de la psychologie cognitive.* Montréal : Les Éditions Logiques.

TIMMERS, L. & CLAEYS, W. (1990). The generality of mental addition models : Simple and complex addition in a decision-production task. *Memory & Cognition, 18*, 310-320.

TOLLEFSRUD, A., CAMPBELL, R.L., STARKEY, P. & COOPER, R.G. (1991). Conservation du nombre : distinguer les solutions par quantification des solutions par opérateurs. In J. BIDEAUD, C. MELJAC & J.-P. FISCHER (Eds), *Les chemins du nombre* (pp.183-210). Lille : PUL.

TREMBLAY, G. & DEMERS, C. (1988). *Référenciel en communication écrite : Des outils pratiques pour développer l'habileté à écrire chez des élèves du primaire.* Rimouski, Québec : Éditions l'Artichaut.

TREMBLAY, G. & DEMERS, C. (1990). Au primaire : Écrire et s'autocorriger. *Vie Pédagogique, 64*, 19-26.

TREMBLAY, G. & SCALLON, G. (1993). *Didactique des projets de lecture, Tome 1 : L'instrumentation.* Rimouski, Québec : Éditions l'Artichaut.

TUDDENHAM, R. D. (1971). Theoretical regularities and individual idiosyncrasies. In *Measurement and Piaget.* New York : McGraw-Hill.

VAN DAAL, V.H.P. & REITSMA, P. (1990). Effects of independent word practice with segmented and whole-word sound feedback in disabled readers. *Journal of Research in Reading, 13,* 133-148.

VAN DAAL, V.H.P. & REITSMA, P. (1993). The use of speech feedback by normal and disabled readers in computer-based reading practice. *Reading and Writing : An interdisciplinary Journal, 5,* 243-259.

VAN DAAL, V.H.P. & VAN DER LEIJ, A. (1992). Computer-based reading and spelling practice for children with learning disabilities. *Journal of Learning Disabilities, 25,* 186-195.

VANLEHN, K. (1987). Learning one subprocedure per lesson. *Artificial Intelligence, 31*, 1-40.

VARELA, F. (1989). *Connaître. Les sciences cognitives, tendances et perspectives.* Paris : Seuil.

VELLUTINO, F. (1979). *Dyslexia : Theory and research.* Cambridge, MA : MIT Press.

VENEZKY, R.L. & SURAJ, M. (1993). Automatic syllabication for an on-line reading tutor. *Behavior Research Methods, Instruments, & Computers, 25,* 67-75.

VERGNAUD, G. (1991), L'appropriation du concept de nombre : un processus de longue haleine. In J. BIDEAUD, C. MELJAC ET J.-P. FISCHER (Eds.), *Les chemins du nombre* (pp.271-282). Lille : PUL.

VERONIS, J. (1986). Etude quantitative sur le système graphique et phono-graphique du français. *Cahiers de Psychologie Cognitive, 6*, 501-531.

VILETTE, B. (1993), Etude expérimentale sur la conservation du nombre, *Cahiers du Laboratoire de Pédagogie expérimentale de l'Université de Louvain, 7*, 91-103.

VILETTE, B. (1994). Des processus de quantification à la cardinalité. *L'année Psychologique, 94*, 25-44.

VYGOTSKY, L.S. (1984). *Pensée et langage.* Paris : Editions Sociales (1ère éd. 1934).

WAGNER, R. & TORGESEN, N.J. (1987). The nature of phonological processing and its causal role in acquisition of reading skills. *Psychological Review, 101*, 192-212.

WAGNER, R.K. & STERNBERG, R.J. (1984). Alternative conceptions of intelligence and their implications for education. *Review of Educational Research, 54*, 179-223.

WEST, R. F. & STANOVICH, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Development, 49*, 717-727.

WEST, R. F. & STANOVICH, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology : Learning, Memory and Cognition, 8*, 385-399.

WIDAMAN, K.F. & LITTLE, T.D. (1992). The development of skill in mental arithmetic : An individual differences perspective. In J.I.D. CAMPBELL (Ed.), *The nature and origins of mathematical skills* (pp.189-253). Amsterdam : Elsevier.

WIDAMAN, K.F., LITTLE, T.D., GEARY, D.C. & CORMIER, P. (1992). Individual differences in the development of skill in mental addition : Internal and external validation of chronometric models. *Learning and Individual Differences, 4*, 167-213.

WINKELMAN, J.H. & SCHMIDT, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology, 102*, 734-736.

WISE, B.W. (1988). Word segmentation in computerized reading instruction. *Dissertation Abstracts International, 48*, 2478-2479B.

WISE, B.W. (1992). Whole words and decoding for short-term learning : Comparisons on a «talking-computer» system. *Journal of Experimental Child Psychology, 54,* 147-167.

WISE, B.W. & OLSON, R.K. (1992). How poor readers and spellers use interactive speech in a computerized spelling program. *Reading and Writing : An interdisciplinary Journal, 4,* 145-163.

WISE, B.W., OLSON, R.K., ANSTETT, M., ANDREWS, L., TERJAK, M., SCHNEIDER, V., KOSTUCH, J. & KRIHO, L. (1989). Implementing a long-term computerized remedial reading program with synthetic speech feedback : Hardware, software, and real-world issues. *Behavior Research Methods, Instruments, & Computers, 21,* 173-180.

WISE, B.W., OLSON, R.K. & TREIMAN, R. (1990). Sub-syllabic units in computerized reading instruction : Onset-rime versus post-vowel segmentation. *Journal of Experimental Child Psychology, 49,* 1-19.

WOLFS, J.-L. (1991). *Analyse de l'anticipation de questions comme indicateur métacognitif.* Thèse de doctorat non publiée, Université libre de Bruxelles, Bruxelles.

WOLFS, J.-L. (1992). Etude des relations entre performances cognitives et métacognition : revue de la littérature. *Recherche en Education , 10,* 15-23.

WOODS, S. S., RESNICK, L. B., & GROEN, G. J. (1975). Experimental test of five process models for subtraction. *Journal of Educational Psychology, 67,* 17-21.

YOUNG-LOVERIDGE, J.M. (1987). Learning mathematics. *British Journal of Developmental Psychology, 5,* 155-167.

ZAGAR, D. (1992a). L'approche cognitive de la lecture : de l'accès au lexique au calcul syntaxique. In M. FAYOL, J.E. GOMBERT, P. LECOCQ, L. SPRENGER-CHAROLLES & D. ZAGAR (Eds.), *Psychologie Cognitive de la Lecture* (pp.15-72). Paris : P.U.F.

ZAGAR, D. (1992b). L'évaluation de la lecture : comment construire un instrument d'évaluation diagnostique ? In P. LECOCQ (Ed.), *La lecture : processus, apprentissage, troubles.* Lille : PUL.

ZAGAR, D., JOURDAIN, C. & LÉTÉ, B. (1993). *ECCLA : Evaluation-diagnostic des Capacités Cognitives du Lecteur Adulte.* Paris : Centre National de Documentation Pédagogique.

ZAGAR, D., JOURDAIN, C. & LÉTÉ, B. (à paraître). Le diagnostiv cognitif des capacités de lecture : le logiciel ECCLA. *Revue Française de Pédagogique.*

ZAKAY, D. (1992). The influence of computerized feedback on overconfidence in knowledge. *Behavior and Information Technology, 11,* 329-333.

ZBRODOFF, N.J. & LOGAN, G.D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology : Learning, Memory, and Cognition, 16,* 83-97.

ZENTALL, S.S. (1990). Fact-retrieval automatization and math problem solving by learning disabled, attention-disordered, and normal adolescents. *Journal of Educational Psychology, 82,* 856-865.

ZENTALL, S.S. & SMITH, Y.N. (1993). Mathematical performance and behavior of children with hyperactivity with and without coexisting aggression. *Behavioral Research Therapy, 31,* 701-710.

Index des tableaux

Première partie

Deuxième partie

Chapitre 9

Index des figures

PREMIÈRE PARTIE

TROISIÈME PARTIE

Chapitre 7

QUATRIÈME PARTIE

Chapitre 9

Contributeurs

BIDEAUD Jacqueline, Université Charles de Gaulle - Lille 3, Département de Psychologie, B.P. 149, 59653 Villeneuve d'Ascq Cedex, France

FISCHER Jean-Paul, IUFM de Lorraine, 16, rue de la Victoire, 57158 Montigny-lès-Metz, France.

GRÉGOIRE Jacques, Université Catholique de Louvain, Faculté de Psychologie et des Sciences de l'Education, 10, Place du Cardinal Mercier, 1348 Louvain-la-Neuve, Belgique.

JOURDAIN Christine, Laboratoire d'Etudes des Apprentissages et du Développement - C.N.R.S., Faculté des Sciences, 6, Bd Gabriel, 21000 Dijon, France.

LAVEAULT Dany, Université d'Ottawa, Faculté d'Education, C.P. 450, succ. A, Ottawa (Ontario) K1N 6N5, Canada.

LÉTÉ Bernard, Institut National de Recherche Pédagogique, 29, rue d'Ulm, 75230 Paris, France.

RETSCHITZKI Jean, Université de Fribourg, Institut de Psychologie, Route de Fougères, 1701 Fribourg, Suisse.

SCALLON Gérard, Université Laval, Faculté des Sciences de l'Education, Département Mesure et Evaluation, Sainte Foy G1K 7P4, Canada.

WOLFS José-Luis, Université Libre de Bruxelles, Département de Didactique Expérimentale, C.P. 186, 50, Avenue F.D. Roosevelt, 1050 Bruxelles, Belgique.

ZAGAR Daniel, Laboratoire d'Etudes des Apprentissages et du Développement - C.N.R.S., Faculté des Sciences, 6, Bd Gabriel, 21000 Dijon, France.